MICROFLUIDIC FRONT-END TECHNOLOGIES FOR PROTEIN ELECTROSPRAY MASS SPECTROMETRY

THÈSE N^{O} 3462 (2006)

PRÉSENTÉE LE 17 MARS 2006

À LA FACULTÉ SCIENCES DE BASE

Laboratoire d'électrochimie physique et analytique

SECTION DE CHIMIE ET GÉNIE CHIMIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Niels LION

DEA de génie biologique et médical, Université de Paris XII, France
et de nationalité française

acceptée sur proposition du jury:

Prof. K. Johnsson, président du jury
Prof. H. Girault, directeur de thèse
Prof. D. Hochstrasser, rapporteur
Prof. H. Lashuel, rapporteur
Prof. J.-D. Tissot, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Lausanne, EPFL
2006

Le réel n'est pas vrai. Il se contente d'être. Et nous construisons une vérité autour de lui, puis une autre, comme un ornement; non pas de façon arbitraire bien sûr, mais en vue de certains objectifs. Pour atteindre une maîtrise objectivable de la Nature.

Henri Atlan.
A tort et à raison.

Remerciements

En premier lieu, je tiens à adresser mes plus chaleureux remerciements à Hubert Girault, qui m'a accueilli au sein de son laboratoire et m'a donné l'opportunité d'entreprendre un travail de thèse, à un moment particulièrement difficile de ma vie où la poursuite d'une activité de recherche n'allait pas de soi. Qu'il soit également remercié pour la constante confiance qu'il m'a témoignée, et pour l'autonomie qu'il m'a accordée tout au long de ces quatre années de thèse. Son enthousiasme et sa créativité resteront un exemple pour moi.

Ces années de labeur ont aussi été ponctuées de deux formidables rencontres humaines, de ces merveilleuses surprises que la vie réserve : j'adresse mes plus affectueux remerciements à Tatiana, qui a été, est et restera plus qu'une amie, mieux qu'une sœur, qui a toujours été là pour partager les bons moments et m'a apporté tout son soutien dans les moments difficiles. Mes remerciements vont aussi à Laura, bien au-delà des travaux menés en commun, pour tous les merveilleux moments passés à Münster, pour toutes les discussions culturelles jusqu'au cœur de la nuit, pour sa gentillesse, sa bienveillance et ses encouragements constants.

Un travail de recherche ne s'accomplit jamais seul. Je tiens tout d'abord à remercier Joël, pour son exemple de créativité, et d'enthousiasme, ainsi que pour sa générosité à faire partager son expérience, ses contacts, à ceux avec qui il travaille, ainsi que pour prouver chaque jour que la réussite et l'accomplissement ne se trouvent pas forcément au bout du chemin le plus clair et le plus traditionnel. Je remercie également profondément tous ceux avec qui j'ai pu travailler au cours de ces années, qui ont partagé avec moi leur savoir, leur savoir-faire, leur gentillesse et leur bonne humeur : Tatiana, Loïc, Michel, Trang, Andrea,

Maurizio, Xiaoxia, Sorina, Jacques, Zhiyong, Momo, Kristof, Jifeng… ainsi qu'André Fattet et tous les membres de l'atelier d'électronique pour leur gentillesse, leur disponibilité et leur grande compétence. J'aimerais également adresser des remerciements tout particuliers à Jean-Olivier Gellon, qui m'a non seulement épaulé dans mon travail, mais qui a également su m'apporter son soutien et son amitié dans les moments difficiles.

Ces années ont également été ponctuées d'activités d'enseignement, activités bien plus enrichissantes encore sur le plan humain que sur le plan scientifique. Je dois beaucoup aux étudiants qui m'ont subi, et j'espère modestement leur avoir apporté autant qu'ils m'ont enrichi. Qu'ils soient tous remerciés collectivement et chaleureusement.

Ce travail n'aurait sans doute pas été le même sans les bienveillants conseils de tous les chercheurs avec qui j'ai pu collaborer au cours de ces quatre ans ; j'adresse mes plus respectueux remerciements au Prof. Dr. Jasna Peter-Katalinic, au Dr. Alina Zamfir, et à Martin Froesch, de l'université de Münster, aux Drs Alan Marshall, Mark Emmett, Logan MacKay et Tukiet Lam, du National High Magnetic Field Laboratory, Tallahassee, aux Pr Michael Przybylski, aux Drs Nikholai Youhnovski et Eugen Damoc, de l'université de Konstanz, au Dr Voyksner de LCMS Limited, au Dr Ken Thomer, du NIEHS. Que tous ceux que j'ai pu oublier veuillent bien par avance accepter mes plus plates excuses.

Ce manuscrit n'aurait pas atteint sa forme finale sans les commentaires pertinents des membres de mon jury de thèse, les Prs Hochstrasser, Tissot, et Lashuel. Qu'ils soient remerciés pour leur lecture critique de ce manuscrit. Je remercie le Pr. Kai Johnsson pour avoir présidé mon jury de thèse.

La fin d'une thèse est aussi l'occasion de se retourner sur les jeunes années de formation ; c'est avec émotion que j'adresse mes chaleureux remerciements aux trois autres membres de notre joyeux quatuor de Saclay, à qui je dois beaucoup plus dans ma formation

scientifique que je n'aurais pu l'imaginer alors ; Frédéric, Richard et Agathe, je vous adresse mes plus amicales pensées.

J'aimerais également témoigner de toute ma gratitude et mon affection à ma « famille suisse » qui m'a accueilli à bras ouvert et m'a permis d'être un expatrié heureux : merci donc à Tatiana et Olivier, Véronique et Fred, puis Mattéo, puis Erin, Agnès et Beat, puis Anoushka, Philou et Simona, Nao, Lucy, Manu, PF et Justine…

Enfin, j'aimerais remercier celle qui a vécu à mes côtés les derniers mois de cette thèse, probablement les plus stressants, qui m'a supporté et encouragé, m'a permis de garder les pieds sur terre et m'a obligé à ne pas m'enfermer (trop) dans le travail. Qui a fait preuve d'énormément de patience, de compréhension et d'amour. Que je ne saurais remercier comme elle le mérite, sinon en lui témoignant de tout mon amour.

Résumé

L'objet de ce travail est de tester et caractériser un système microfluidique d'électronébulisation pour l'analyse de biomolécules par spectrométrie de masse, et plus particulièrement dans un contexte d'analyse protéomique. Etant donné que ce travail est essentiellement analytique et technologique, le Chapitre 1 décrit le contexte épistémologique qui a vu l'émergence de la protéomique et de la biologie systémique. Il est démontré que ce nouveau paradigme n'est pas apparu *ex nihilo*, mais est le résultat de basculements épistémologiques majeurs en médecine, physique et ingénierie. Par ailleurs, ce changement paradigmatique s'inscrit dans la longue histoire des luttes entre conceptions holistes et programmes réductionnistes en biologie et médecine. Finalement, le potentiel biomédical de la biologie systémique est discuté.

Le Chapitre 2 présente l'état de l'art dans le couplage de systèmes microfluidiques à la spectrométrie de masse par électronébulisation, tant du point de vue technologique qu'applicatif. Les Chapitres 3 et 4 présentent la caractérisation du nouveau micro-électronébulisateur polymère pour l'analyse de peptides, protéines et glycoconjugués sur différents instruments. En particulier, une comparaison détaillée est faite avec des nano-électronébulisateurs capillaires pour préciser les performances du nouveau système en termes de sensibilité, de stabilité, de gamme de débits applicables et de compatibilité avec les solvants organiques.

Le Chapitre 5 introduit un micro-électronébulisateur fonctionnalisé pour la purification d'échantillons. En résumé, une membrane hydrophobe de fluoride de polyvinylidène est interfacée à l'entrée du micro-électronébulisateur polymère pour servir de phase solide de capture de composés hydrophobes ; une fois capturés, ces composés

peuvent être lavés, puis élués par la solution d'électronébulisation qui contient des solvants organiques. Ce procédé est également évalué lorsque l'échantillon de départ contient des composés courants en préparation d'échantillons protéiques, tels que chaotropes, agents réducteurs, ou détergents.

Le Chapitre 6 décrit un développement plus important du système microfluidique lui-même, qui permet d'électronébuliser deux solutions en même temps ; de par la conception du microsystème, les deux solutions ne se mélangent que dans le cône de Taylor, où l'électronébulisation a lieu, et les débits des deux solutions peuvent être contrôlés indépendamment. Cette particularité unique permet d'analyser des échantillons purement aqueux, en intégrant un minimum de solvant organique pour permettre la désolvatation et l'ionisation dans de bonnes conditions. De plus, des protéines peuvent être analysées en conditions natives, comme l'atteste leur distribution de charge dans le spectre de masse. En jouant sur le débit de solvant organique injecté directement dans le cône de Taylor, on peut induire une dénaturation en ligne de la protéine analysée, et suivre cette dénaturation par l'évolution de son état de charge dans le spectre de masse. Le potentiel de cette technologie pour la caractérisation de la stabilité thermodynamique de protéines est discuté.

Finalement, le Chapitre 7 présente une évaluation *in silico* de l'intérêt de deux technologies développées au laboratoire : l'électrophorèse Off-Gel pour le fractionnement isoélectrique de peptides et protéines, et le comptage en ligne des cystéines contenues dans des peptides lors de leur analyse par spectrométrie de masse, pour l'analyse protéomique rapide. Si l'on recherche dans un protéome entier digéré par protéolyse les peptides uniques, la mesure de leur masse seule, même avec un spectromètre à très haute résolution et à très

haute précision en masse, ne permet pas d'obtenir des identifications de protéines utiles. Le but de ces simulations est d'évaluer dans quelle mesure l'ajout de l'information sur le point isoélectrique de tous les peptides, ainsi que de leur contenu en cystéines, à la mesure de leur masse, permet d'envisager des mesures protéomiques haut-débit.

MOTS-CLEFS : protéomique, spectrométrie de masse, microfluidique, biochimie analytique

Abstract

In the present work, a polymer microfabricated microsprayer is characterised and tested for the mass spectrometric analysis of biomolecules, in particular in the context of protein mass spectrometry and proteomics. As the core of this work deals with analytical sciences and device development, Chapter 1 puts this work in perspective with the emergence of proteomics and more generally systems biology. In particular, major epistemological concepts that allowed the emergence of these new disciplines among life science research are reviewed. It is shown that the new paradigm of integrative biology did not emerge *ex nihilo*, but was preceded by major epistemological shifts in medicine, physics, and engineering. As such, the rise of systems biology appears as a new episode in the balance between reductionist and holistic approaches in the history of biology and medicine, and its biomedical promises are discussed.

Chapter 2 reviews the state of the art in the hyphenation of microfluidic devices with electrospray mass spectrometry, both from a technical and applicative standpoint. Chapter 3 and 4 present the characterisation of the new polymer microspray for the analysis of peptides, proteins and glycoconjugates when coupled with various electrospray ion sources and instruments. In particular, a detailed comparison with classical pulled nanospray capillaries is performed, that established the performances in terms of sensitivity, stability and applicable ranges of flow rates and solvents.

Chapter 5 introduces a functionalised polymer microspray for online sample clean-up. Basically, a hydrophobic polyvinylidene fluoride membrane is interfaced at the inlet of the polymer microspray to serve as a capture solid-phase of hydrophobic compounds; once captured, they can be cleaned up, and further eluted by the spray solution containing

organic solvents. The process is further investigated in the presence of compounds usually used in protein sample preparation, such as chaotropes, reducing agents or detergents.

Further microchip developments include the introduction of a dual channel microsprayer in Chapter 6, that allows to spray two solutions at the same time; given the microchip design, the two solutions are mixed only in the Taylor cone, where the electrospray is generated, and their respective flow rates can be controlled independently. This unique feature allows to analyse pure aqueous samples with integration of minimum amounts of organic sheath flow to promote desolvation and ionisation. Moreover, when proteins are analysed, it allows to measure them in their folded state, as demonstrated by their charge state distribution in the mass spectrum. Further tuning of the organic sheath flow allows to denature the protein within the Taylor cone, which can be monitored by the evolution of the protein charge state distribution in the mass spectrum. Potential of this technology for protein thermodynamic stability measurement is discussed.

Lastly, Chapter 7 presents *in silico* evaluation of two technologies developed in the laboratory, namely Off-Gel electrophoresis for isoelectric fractionation of proteins and peptides mixtures, and online counting of cysteine residues within peptides during their analysis by electrospray mass spectrometry, to provide useful information to speed-up proteome profiling: if one wants to seek within a whole digested proteomes for peptides that are unique, measurement of their mass alone is hardly sufficient to perform useful protein identification, even with very high resolution, high mass accuracy instruments. The purpose of these simulations is to estimate how much information peptide isoelectric point and number of cysteines within each peptide provide in terms of number of unique peptides

(and hence unambiguous peptide identification without MS/MS). Applicability of this strategy for high-throughput proteome profiling and its limitations are further discussed.

KEYWORDS: proteomics, mass spectrometry, microfluidics, analytical biochemistry

Table of content

Chapter 1. Introduction

In the last two decades, mass spectrometry has become a central analytical technique in biochemistry and especially in protein analysis, due to its capability to provide not only mass information, but also identification when combined with enzymatic procedures and database search, as well as structural information. It is now established as the technique of choice for protein identification from biological samples. This emergence stems back from the development of two soft ionisation techniques, electrospray ionisation (ESI) [1] and matrix-assisted laser desorption/ionisation (MALDI) [2-4] during the late eighties, that allow gentle ionisation of large biomolecules. The importance of these new ionisation techniques beyond the field of analytical chemistry was recently recognised by the attribution of the chemistry Nobel Prize to John Fenn [5] and Koichi Tanaka [6] in 2002. The advent of ESI and MALDI ion sources, together with other developments in mass spectrometry, have largely contributed to the emergence of proteomics as a new field, by allowing researchers to massively and rapidly identify proteins, with unprecedented precision and comprehensiveness. However, electrospray ionisation sources suffer from technological drawbacks: the most common technology consists in manually pulling fused silica capillary, to produce capillaries with small outlets that are favourable for sample delivery to mass spectrometers through electrospray ionisation. This manual procedure is tedious and requires trained operators, have a limited success yield because of capillary breakage during the procedure, and usually results in poorly reproducible capillaries.

In the late nineties, Ramsey's [7] and Karger's [8, 9] groups introduced new electrospray ionisation sources that were not based on pulled capillaries, but on microfluidic chips produced with technologies imported from the microelectronic industry. The basic underlying idea was to use technologies that could potentially produce reproducible electrospray ionisation sources with dimensions in the low micron range, on an industrial scale. The work presented in this manuscript follows the approach pioneered by Ramsey and Karger, i.e. the development of microfluidic electrospray ionisation sources in a proteomic context: chapter 2 reviews the state of the art in the coupling of microfluidic electrospray ionisation sources with mass spectrometers [10, 11], chapter 3 describes the development of polymer-based microfluidic electrospray sources and their characterisation [12-14], chapter 4 reports their application in the particular field of glycoconjugate analysis [15-17], chapter 5 introduces functionalised microchips for integrated sample desalting prior to electrospray ionisation [18, 19], chapter 6 provides some perspectives for the use of a dual channel microfluidic electrospray ion source, and chapter 7 briefly describes some bioinformatic tools developed to investigate the integration of various analytical technologies developed in our laboratory into a whole proteomic analysis platform.

As the work described in this thesis is mainly technologically oriented, it is worth giving an overview of the scientific context in which it has been carried out.

1. What is proteomics?

The term "proteome" was first introduced by Marc Wilkins at the Sienna meeting in 1994, basically to replace the expression "all proteins expressed by a genome, cell or tissue". Since then, proteome analysis, or proteomics, has somehow gained in scope: whereas no clear consensus definition exists, proteomics can be defined as the "identification of all proteins expressed by a genome, cell or tissue, their quantification, and

determination of their interactions". Whereas the definition of proteomics given above does not refer to any particular technique, proteomics is often understood as the set of techniques used in proteomics, such as two-dimensional gel electrophoresis (2D-GE), multidimensional liquid chromatography (LC), MALDI-Time-Of-Flight-Mass Spectrometry (MALDI-TOF-MS), LC-MS/MS…

Basically, the goal of proteomic projects are of two different kinds: on one hand, the purpose is to identify the whole set of proteins present in a given sample, which we propose to call *discovery proteomics*, in the case where this set of proteins has never been observed, or *profiling proteomics*, in the case where this set of proteins is investigated in a particular case (such as a sample from a disease patient, or a cell line in given conditions of growth). To a certain extent it is similar to genome sequencing, whose primary goal was to provide biologists with the list of genes and their sequence (whatever the "political" claims of all the genome sequencing projects were, it was clear from the beginning that the main output of these projects would be "reference data" for biologists, and that biomedical discoveries would be limited to some monogenic diseases, at least in the first decades after the release of genome sequences). Subfields of discovery and profiling proteomics are post-translational mapping (identification of phosphorylation, glycosylation, ubiquitination…. sites in the identified proteins, or identification of phosphorylated or glycosylated or ubiquinated sub-classes of proteins from a sample), chemical proteomics, whose aim is to identify proteins with a particular enzymatic function, and interaction proteomics, whose goal is to identify the whole set of protein/protein interactions.

The other mainstream in proteome research is the search for disease biomarkers, i.e. proteins or small sets of proteins that are present in much greater or much lower quantity in patients with a given disease compared to healthy subjects, or which present given alterations (such as different post-translational modifications patterns). In a way it is similar

to classical biomedical research, in which typical levels of particular biomolecules (or their modifications) are used to classify individuals between healthy and disease. The purpose of biomarker research is first to develop diagnosis or prognosis tools. In a second step, it is believed that if a particular protein is particularly over- or under-expressed in a disease patient, this protein may play a role in the disease process, and thus may become a target for therapy (see below for a discussion of the conceptual validity of this approach).

From a technical point of view, there are two main approaches in proteomics. One is based on 2D-GE separation of all proteins present in a sample: proteins are first separated in an immobilised pH gradient (IPG) strip according to their isoelectric points (in denaturing conditions); once focused, detergent is added so that focused neutral proteins get capped with charge surfactant molecules and acquire a charge proportional to their length (and thus to their sizes in a first approximation). The IPG is then transferred on top of a polyacrylamide gel, and electrophoresis is run in the second dimension to separate proteins according to their mass; after electrophoresis, the gel is stained for example by Coomassie blue or silver in order to visualise protein spots, as shown in Figure 1. In order to identify proteins, spots are excised from the gel, proteins are in-gel digested by trypsin and the resulted peptides are spotted onto a MALDI plate in order to measure their mass by MALDI-TOF MS. Tryptic peptide masses constitute a fingerprint of the protein that can then be identified through interrogation of proteins (e.g. SwissProt) or DNA (e.g. Trembl) databases [20, 21]. Though 2D-GE suffers from a number of caveats (in particular in coping with the high diversity of protein physico-chemical properties such as size, hydrophobicity or charge) [22], it is still the workhorse for proteome research [23].

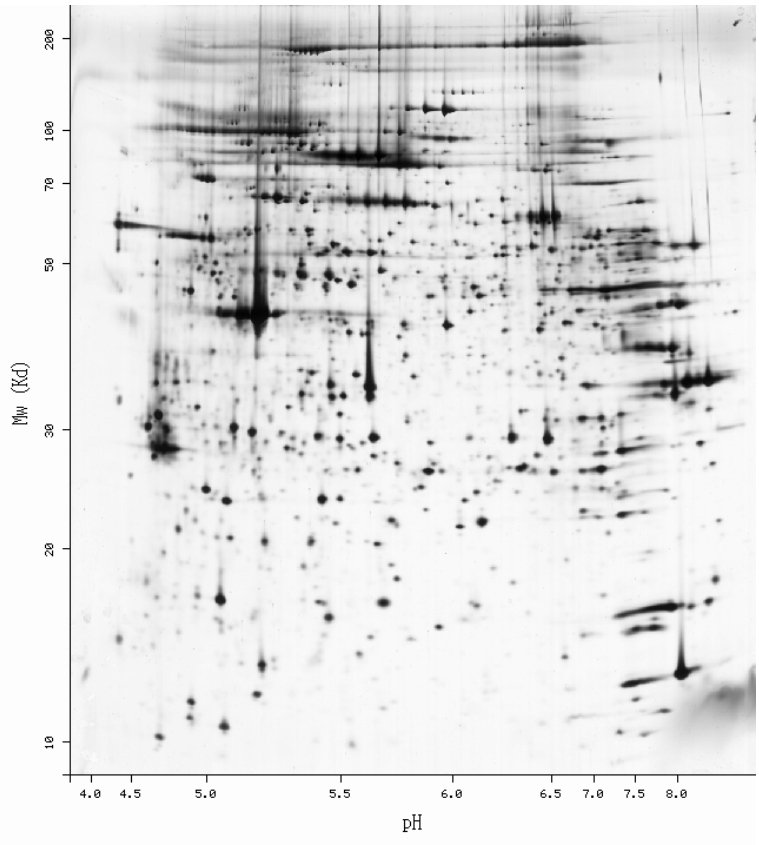


Figure 1. 2D gel map of human platelet, as retrieved from Swiss2Dpage (August 2005). This SWISS-2DPAGE entry is copyright the Swiss Institute of Bioinformatics.

The second analytical approach used in proteomics is a gel-free, shotgun apporach: the whole set of proteins is first digested by trypsin to produce a peptide mixture; the basic idea is that peptides have much more homogeneous physico-chemical properties than proteins (in terms of size, hydrophobicity and charge) and thus can be handled more efficiently by routine analytical procedures. The counterpart of this gain in properties homogeneity is that each protein typically gives a few tens to a few hundreds of peptides; so instead of separating a few hundred or a few thousand proteins, the challenge is to separate up to a few hundred thousand peptides before delivering them to the mass spectrometer. This approach has thus relied on the development of new and powerful liquid-phase separation tools, such as the dual phase microcolumns introduced by Wolters *et al* [24] or the solid phase capture of cysteine containing peptides prior to LC-MS/MS developed by

Gygi *et al* [25]. In the shotgun approach, liquid chromatography is directly interfaced to electrospray mass spectrometers in order to analyse the separated peptides online. Obviously, the measurement of peptide masses alone (as done in the peptide mass fingerprinting approach) is not sufficient to identify proteins through database interrogation from such complex mixtures, because peptides originating from a single protein can span over the whole chromatographic run (or conversely each peptide eluting from liquid chromatography at a given time can originate from a different protein). In fact, online peptide sequencing by tandem mass spectrometry is necessary to produce partial sequence information; database interrogation is then based on peptide masses and on peptide partial sequence information [26]. This need for peptide sequencing at the pace of the LC separation imposes great constraints on the sampling rate and MS/MS capacity of mass spectrometers. New instruments have been recently introduced to cope with this demand, such as the linear ion trap [27]. Figure 2 shows the schematic workflow for a proteomic shotgun experiment:

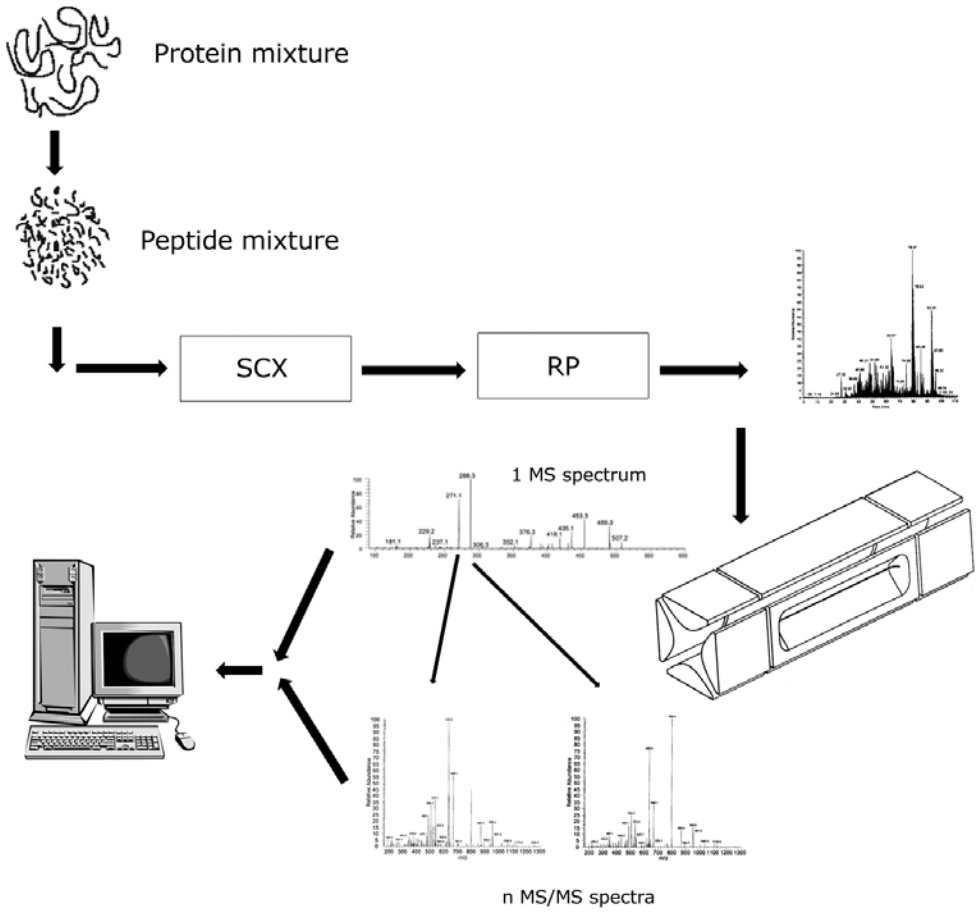


Figure 2. Shotgun proteomics workflow: the whole protein mixture is digested; the resulting peptide mixture is subjected to two-dimensional liquid chromatography (usually strong cation exchange (SCX) and reverse phase (RP)); the mass of eluting peptides are measured by the mass spectrometer, and in each spectrum, typically the three most intense peaks are automatically subjected to MS/MS. The collection of MS and MS/MS spectra is then submitted to database interrogation for protein identification.

The typical result of such experiments (whether by 2D-GE and MALDI-TOF-MS or by LC-ESI-MS/MS) is a list of identified proteins, which is usually completed by a repartition of the identified proteins by function or localisation, as shown in Figure 3.

Cellular designation	Annotation	Protein name	Spot No.	MALDI-MS			LC-MS/MS
				Mascot	Profound	Seq Cov	Mascot
Plasma	SP_P01876	Ig α-1 chain C region	71	47	2.1	26	
Plasma	SP_P01877	Ig α-2 chain C region	71	43	2	13.5	213
Plasma	SP_P02647	Apilipoprotein A-I	3	49	2.4	29.6	683
Plasma	SP_P02649	Apolipoprotein E	53	42	2.4	26	674
Plasma	SP_P02652	Apolipoprotein A-II	12				99
Plasma	SP_P02655	Apolipoprotein C-II	58				95
Plasma	SP_P02656	Apolipoprotein C-III					92
Plasma	SP_P02735	Serum amyloid A protein	19	94	2	56	
Plasma	SP_P02741	C-reactive protein		36	2.2	24	
Plasma	SP_P02743	Serum amyloid P component		36	1.9	30	
Plasma	SP_P02748	Complement component C9	55	36	2.3	18	
Plasma	SP_P02749	β-2 glycoprotein I	34				232
Plasma	SP_P02750	Leucine-rich α-2 glycoprotein	31	44	2.4	26.9	277
Plasma	SP_P02751	Fibronectin	64				358
Plasma	SP_P02753	Plasma retinol-binding protein					201
Plasma	SP_P02760	α-1 Microglobulin		37	2.3	21.6	175
Plasma	SP_P02765	α-2-HS glycoprotein		59	2	17.7	
Plasma	SP_P02766	Transthyretin	46	48	2.1	49	
Plasma	SP_P02768	Albumin		49	2.2	11.8	449
Plasma	SP_P02774	Vitamin D-binding protein	29	48	2.4	16.7	368
Plasma	SP_P02787	Transferrin	30	47	2.1	16.3	591
Plasma	SP_P02790	Hemopexin	36				102
Plasma	SP_P03952	Plasma kallikrein					176
Plasma	SP_P04003	Complement C4b-binding protein, alpha chain	54				226
Plasma	SP_P04004	Vitronectin		39	2.4	17	
Plasma	SP_P04070	Vitamin K-dependent protein C		36	2.9	16.3	
Plasma	SP_P04114	Apolipoprotein B-100					451

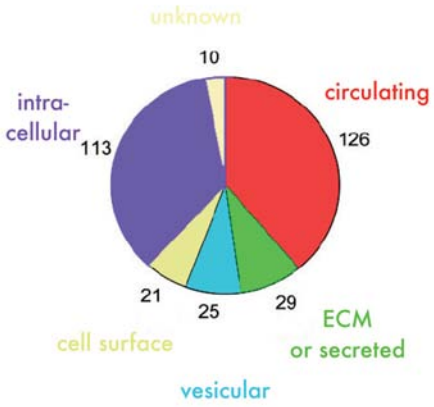


Figure 3. Left: top of the list of proteins identified in plasma (out of 325). Results consist in the protein localisation, the SwissProt accession number, and identification parameters. Right: Localisation of the identified proteins. From [28].

The work presented in the following chapters focuses on the design and characterisation of microfabricated electrospray emitters. As such, its scope in a whole proteomic platform concerns only the hyphenation of liquid phase analytical processes to electrospray mass spectrometry. It is thus worth giving an overview of the concepts at work in proteomics. The following sections trace back the origins of the paradigms into which proteomics emerged recently and some aspects of the history of biological research whose contribution are often underrated in the appearance of proteomics and more generally systems biology.

2. A brief and partial history of life science research

Proteomics is considered by many as a new *paradigm* [29][1]. In a way, proteomics is not a new paradigm as defined by Thomas Kuhn, because its status is highly debated: some consider it as a simple continuation of the genome revolution, some others as a revival of classical physiology [30]. In this section, I will try to show that proteomics, as a part of systems biology, truly consists in a new paradigm, in the sense introduced by Thomas Kuhn: it is a scientific field with its own subjects of study, questions, tools, instruments, methodologies, laboratories, conferences, networks… However, the conceptual framework that allowed its appearance is often ignored or disregarded. Most scientists currently working in life sciences and analytical chemistry have been educated in the context of molecular reductionism, whose methodology is to study all the parts of a living system within their most intimate details in order to understand the functioning of the whole. This research program is embodied in life sciences through the development of specialised disciplines such as endocrinology, enzymology, protein biochemistry, molecular biology…

On the contrary, classical physiology does not aim at describing every molecule present in a biological sample; physiology aims at a global understanding of living organisms through analysis of their main functions. Through history, physiology has developed into a molecular discipline based on the reductionism dogma (see below). At first sight, systems biology positions itself somewhere between these two poles: on one hand it aims at collecting molecular information on a large scale, which is a purely reductionist scientific enterprise; on the other hand it aims at integrating this molecular information into

[1] A paradigm is here understood as defined by Thomas Kuhn in *The structure of scientific revolutions*, i.e. as "a set of theories, attitudes, and procedures that scientists do not question".

valuable knowledge to better understand living beings. Whereas for most scientists it may appear as a similar goal, it fundamentally differs in the fact that for systems biologists, molecular determinism cannot explain everything. A classical problem that reductionism has always been unable to decipher is how a full organism with head, arms, antennas, stomach, tail… can originate from a single cell dividing into two, four, eight… undifferentiated cells. In a way, the Human Genome Organisation failed to fulfil its biomedical promises (at least the public expectations), because its outputs were expected in the context of molecular determinism: each disease was supposed to be fully explained by one gene or in the worst case a small set of genes, that would account for all pathological aspects. But despite the completion of human genome sequencing, no major breakthrough in the understanding of disease has been obtained so far. Gene knowledge is now being used in medicine as a predisposition diagnosis and treatment orientation and prognosis tool, as is the case in breast and ovarian cancers with the BRCA gene family [31-33]. But very limited progress has been made in the design of new drugs based on genomic information. This limited success, and the claims made by proteomic or systems biology pioneers seriously question the status of these disciplines into the history of life science research. To put it into perspective, main conceptual aspects of the history of life science research are highlighted below.

2.1. From Greeks to mechanistic biology

The Greek physicians Hippocrates (fourth century before Christ) and Galen (second century after Christ) can be considered as the founders of scientific medicine. They eradicated all the magical approaches of disease, and based their approach on empirical healing procedures based on observation. Though very limited by the technical means available at that time, they pioneered anatomy and surgery, elaborated physiological

concepts based on their observations and introduced the rational use of plants as drugs. Their influence was so seminal that in many occidental countries, the Hippocrates oath is a ritual that symbolises the completion of medical training and ability to treat patients, and the word "galenic" has passed into pharmaceutical sciences. During Middle Ages, Arabs took over the Greek and Roman heritage: biology was still tightly associated with medicine, but it was also tainted with philosophy through the contributions of Gnostic philosophers Avicenne and Averroes. Arabs were also the first to try to analyse the living matter in terms of chemistry, and can thus be considered as the pioneers of biochemistry. But scientific biology and medicine made little progress till the seventeenth century mainly because of the lack of experimental tools to study living beings and of a general scientific program. The situation changed during the 17th century, mainly due to the scientific methodology introduced by Descartes.

Descartes is often considered as the father of reductionism, defined as a methodology to study and understand complex phenomena [34]:

> *"I believe that, instead of the multiplicity of rules that comprise logic, I would have enough in the following four, as long as I made a firm and steadfast resolution never to fail to observe them.*
>
> *The first was never to accept anything as true if I did not know clearly that it was so; that is, carefully to avoid prejudice and jumping to conclusions, and to include nothing in my judgements apart from whatever appeared so clearly and distinctly to my mind that I had no opportunity to cast doubt upon it.*
>
> *The second was to subdivide each of the problems I was about to examine into as many parts as would be possible and necessary to resolve them better.*

> *The third was to guide my thoughts in an orderly way by beginning, as if by steps, to knowledge of the most complex, and even by assuming an order among objects in cases where there is no natural order among them.*
>
> *And the final rule was: in all cases, to make such comprehensive enumerations and such general reviews that I was certainly not to omit anything.*
>
> *The long chains of inferences, all of them simple and easy, that geometers normally use to construct their most difficult demonstrations had given me an opportunity to think that all the things that can fall within the scope of human knowledge follow from each other in a similar way, and as long as one avoids accepting something as true which is not so, and as long as one observes the order required to deduce them from each other, there cannot be anything so remote that it cannot be reached nor anything so hidden that it cannot be uncovered." Discourse on method (1637). Descartes.*

Following Descartes, physicians and biologists began to think about living beings in terms of mechanisms made of parts. In this perspective living beings were seen as mechanisms just a little bit more complex than clocks. The first triumph of this mechanistic concept is the description of blood circulation by William Harvey in his treatise *On the motion of the heart* (1628). Heart pumping, arteries and veins were depicted macroscopically in terms of fluid mechanics and anatomy. Other attempts were made during the 17th century to explain other physiological functions by dissecting their parts and rebuild them (practically speaking or by experiment of thought) into true mechanical systems. This approach is best exemplified by Vaucanson's duck:

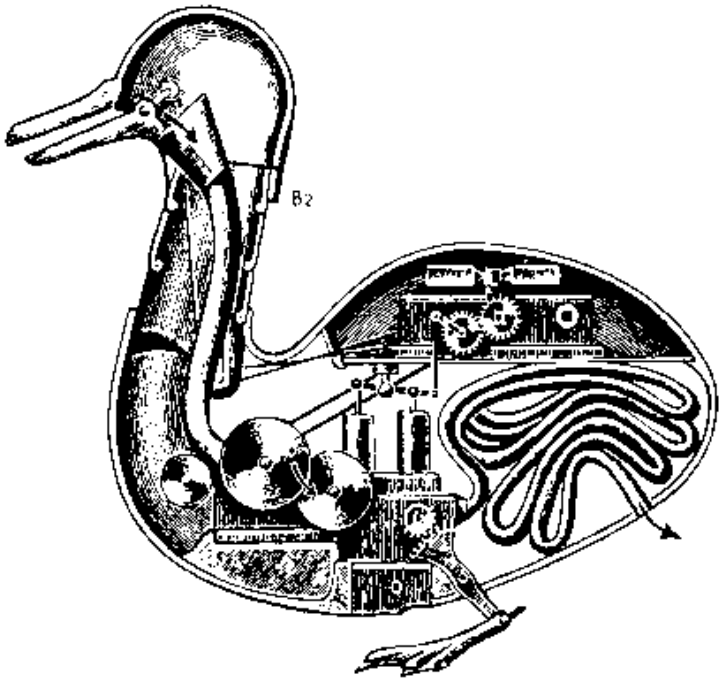


Figure 4. Vaucanson's duck. Automates Museum. Grenoble. France.

Vaucanson built a mechanical duck that was able to ingest, chew and "digest" aliments. Though anecdotic, this automata accounts for the scientific program of the 17th century in life sciences: physiological processes could be understood by dissecting and mimicking them by mechanical devices. This concept was largely based on the fact that the smallest observable biological units were organs and tissues that resemble much mechanical devices such as pumps, springs, or pipes. But this purely mechanistic approach remained largely unsuccessful and main physiological functions such as breathing, digestion, movement… remained unexplained.

Interestingly, at the same time another school of thoughts initiated by Paracelsus attempted to study biological processes in terms of chemistry. They benefited from the Arab alchemists' heritage and built a theory into which life was considered as a pure chemical process, and disease as a disequilibrium of this process. As for Paracelsus, medicine was not only a matter of understanding and curing, but was also based on philosophy and ethics:

> *"Medicine rests upon four pillars—philosophy, astronomy, alchemy, and ethics. The first pillar is the philosophical knowledge of earth and water; the second, astronomy, supplies its full understanding of that which is of fiery and airy nature; the third is an adequate explanation of the properties of all the four elements—that is to say, of the whole cosmos— and an introduction into the art of their transformations; and finally the fourth shows the physician those virtues which must stay with him up until his death, and it should support and complete the three other pillars." Das Buch Paragranum (1529-1530). Paracelsus.*

During all the 17th century, there was a conceptual fight between Descartes' and Paracelsus' heirs: mechanistic view of living beings was opposed to a more holistic view of life as a chemical process. But probably due to the depth of the scientific and philosophical Cartesian program, reductionism (and especially its mechanistic aspect) dominated the 17th century.

The 18th century witnessed a considerable evolution by the development of chemistry and discovery of electricity. Lavoisier discovered the theory of combustion of gases and studied breathing as a particular kind of oxidation processes. At the turn of the century, Galvani discovered that electric current was associated with nerve function, and Volta pioneered electrophysiology by demonstrating that an electric current could induce muscular action (and reversibly). These discoveries of entirely new processes at work in living beings questioned the mechanistic approach that dominated the 17th century. Moreover, as gases and electricity could not be regarded at that time as "cartesian parts" or elemental devices, these new discoveries should have questioned the reductionist program. In practise, the discovery of the role of electric currents in neural and muscular systems and of oxidation in breathing, were perceived as steps that could not yet be reduced to elemental processes at the time, and reductionism (including its mechanistic program) continued to

dominate the occidental life sciences and philosophy. It culminated in the book *Man a machine* (1748) by De La Mettrie who postulated that human beings were pure machines and that even thought and mind would eventually be reduced to a mechanistic description.

The 19th century witnessed an enormous development of biology and medicine, and the true advent of classical life sciences. One revolutionary contribution is the recognition that all living beings are made of cells; first discovered by Hooke in the 18th century through the development of the microscope, only during the 19th century was a true cellular theory elaborated by Schwann:

> *"The elementary parts of all tissues are formed of cells in an analogous, though very diversified manner, so that it may be asserted that there is one universal principle of development for the elementary parts of organisms, however different, and that this principle is the formation of cells". Microscopic researches into the accordance in the structure and growth of animals and plants (1839). Theodor Schwann.*

Though the development of the microscope played a seminal role in the cellular theory, it cannot account solely for the discovery of cells (for a detailed study of the conceptual requisites needed to observe something with a microscope, the reader is referred to reference [35]). The 19th century was marked by the genius of Louis Pasteur who studied bacteria in various contexts; its major contribution was to demonstrate that self-generation does not exist, but that the turbidity that can appear in liquids originates from the growth of originally undetectable microorganisms. The philosophical implications of the impossibility of self-generation were so deep that the controversy about it continued well after the experimental demonstrations carried out by Pasteur. After twenty years of research about bacteria in the context of what would be now called industrial biotechnologies (i.e. the milk, beer and wine industries at that time), Pasteur also demonstrated that many diseases could

be explained by the presence of specific microorganisms in the ill host. By directly relating disease to the proliferation of micro-organic intruders, Pasteur delineated the biomedical research program for the next century, which culminated by the world-wide vaccination campaigns during the 20th century.

The identification of cells as elementary building blocks of living beings, as well as the link established by Pasteur between microorganisms and diseases constituted the most striking successes of the reductionist scientific program, and gave strength to it to an unprecedented level. Historically, these successes corresponded to the rise of scientific positivism that claimed that everything could be understood for the general benefit of Mankind. Unfortunately, this mainstream scientific and philosophical school of thought eclipsed a theory introduced a little before by Claude Bernard, considered as the father of modern physiology. Whereas Bernard was truly faithful to reductionism, its physiology was based on the existence of an inner medium, in which there were organs and tissues. He observed that this inner medium was remarkably constant, independently of the external conditions. He thus postulated that existence and constancy of this inner medium (now called *homeostasis*) was a condition for life, and that disease was associated with disequilibrium of this inner medium. Though the medical methodology introduced by Bernard in the *Introduction to the study of experimental medicine* (1865) established him as a commanding figure in the history of biomedical sciences, his elaborated concept of health and disease was largely disregarded by biologists and physicians till its rediscovery during the 20th century.

One of the major contributions of the 19th century to the history of ideas is the Darwinian theory of evolution. Darwin's theory describes living beings as systems in perpetual evolution and transformation. The world was no more a perfect Newtonian machine created by God, but a system whose components were evolving from one another.

However Darwin's theory did not have any molecular or elementary explanation, and all attempts to describe it in terms of hereditary units failed; for example, Darwin thought that each offspring inherited approximately half of its parents' traits. As evolution was due to random mutations, inheritance mechanisms would then result in quick disappearance of new traits. Ironically, only at the beginning of the 20th century was the work of Gregor Mendel rediscovered by William Bateson, which provided the first elements of a reductionist explanation of mutations and evolution.

The 19th century was the golden era of the reductionist program: the establishment of the cellular theory, the role of bacteria (an later of viruses) in disease, and the subsequent development of vaccination, and the early developments of molecular biology, testified for the strength of the program and eclipsed all other concepts of life, health and disease.

2.2. The century of doubts

The successes obtained during the 19th century by the reductionist approach to life sciences problems marked the early developments of molecular biology; in particular, whereas the molecular structures of heredity were not elucidated yet, the first central dogma of molecular biology became: "one hereditary unit, one trait" [36]. However, following the Mendelian approach of hereditary mechanisms, it was soon discovered that one gene could affect several traits or conversely that one trait could be under the control of several genes. Nevertheless, progress in biochemistry let envisioned a full explanation of life processes in terms of their molecular components:

> *"When the biologist is confronted with the fact that in the organism the parts are so adapted to each other as to give rise to a harmonious whole; and that the organisms are endowed with structures and instincts calculated to prolong their life and perpetuate their race, doubts as to the adequacy of a*

> *purely physicochemical viewpoint in biology may arise. The difficulties besetting the biologist in this problem have been rather increased than diminished by the discovery of the Mendelian heredity, according to which each character is transmitted independently of any other character. Since the number of Mendelian characters in each organism is large, the possibility must be faced that the organism is merely a mosaic of independent hereditary characters. If this be the case the question arises: What moulds these independent characters into a harmonious whole? […] The theory of natural selection invokes neither design or purpose [*as contrary to vitalism*], but it is incomplete since it disregards the physicochemical constitution of living matter about which little was known until recently." The organism as a whole: from a physicochemical viewpoint (1916). Jacques Loeb.*

As summed up by Jacques Loeb, this tension between the early results of Mendelian genetics and the reductionist program was attributed to an insufficient knowledge of molecular components of heredity. Successes of the reductionist program continued to accumulate in the first half of the 20th century, with the discovery that enzymes carry out chemical reactions in living beings and are homologous in many different species (which was in line with the universal elementary units of the reductionist program), and later on that genes were at the origin of the synthesis of enzymes [37]. But biology was still confronted with the lack of information about the intimate structure of the molecules involved in these processes.

The situation dramatically changed during the forties, with the incursion of physicists into biology. Major figures of this trend were Francis Crick, James Watson, Maurice Wilkins, Rosalind Franklin, Linus Pauling, Salvador Luria and Max Delbrück. Beyond biological research, the interest of physicists for biology is best exemplified by the

work of Niels Bohr, who questioned the applicability of the uncertainty principle in biology, and even more by Erwin Schrödinger in his popular book *What is life?*. Schrödinger hypothesised what the molecular basis of heredity should be from very general physical considerations of permanence and information storage. Interestingly, though Schrödinger's goal was to deduce the molecular structure of heredity, his reasoning was based on very general considerations about hereditary processes and a holistic approach of living organisms. But it seems that the only methodological lesson learned from his work was that physical sciences had much to provide to biological reductionism. And major contributions were obtained in the following years thanks to X-ray crystallography: Linus Pauling elucidated the three dimensional structure of proteins [38-40] (Nobel prize in chemistry 1954) and Watson and Crick [41] that of DNA (Nobel prize in medicine and physiology 1962). Within the next decade, the mechanism by which nucleic acids direct protein synthesis were elucidated, and were found to be essentially the same from bacteria to homo sapiens. These results constituted the triumph of the reductionist program in life sciences, and molecular biology became the new frontier of life sciences, with a clear established program:

> *I present them* [the Sequence Hypothesis and the Central Dogma] *here in the hope that others can make similar use of them. Their speculative nature is emphasised by their names. It is an instructive exercise to attempt to build a useful theory without using them. One generally ends in the wilderness.*
>
> The Sequence Hypothesis. *This has been already referred to a number of times. In its simplest form it assumes that the specificity of a piece of nucleic acids is expressed solely by the sequence of its bases, and that this simple sequence is a (simple) code for the amino acid sequence of a particular protein…*

> The Central Dogma. *This states that once "information" has passed into protein* it cannot get out again. *In more detail, the transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible, but transfer from protein to protein, or from protein to nucleic acid is impossible. Information means here the* precise *determination of sequence, either of bases in the nucleic acid or of amino acids residues in the protein. This is by no mean universally held —Sir Macfarlane Burnet, for example, does not subscribe to it— but many workers now think along these lines. As far as I know it has not been* explicitly *stated before". On protein synthesis, symposia of the Society for Experimental Biology: the biological replication of molecules, 1958, 12:152-153. Francis Crick.*

The full paradigm of molecular biology is contained in this quotation from Francis Crick: first, for nucleic acids, the sequence only carries information, and secondly, DNA is at the origin of proteins, and in a more specific way, the Central Dogma of molecular biology (which has been taught to all students since then) can be rephrased as "one gene, one protein". The clearness and the power of these dogmas in light of the reductionist approach successfully drove the booming of molecular biology in the first decades after World War II. Interestingly, Crick called them "Hypothesis" and "Dogma", which implies that at that time they were not established facts supported by solid experimental data and integrated into a coherent theory, but more a research program, or to express it differently, a paradigm to be tested. But the hypothesis was so seducing, and the prospect of explaining every life process through molecular biology seemed so close, that it was rapidly accepted and taken as granted.

However, the molecular biology program eclipsed some major problems in biology, or otherwise stated, moved the centre of the various disciplines, which was to understand

physiological functions, to molecular biology in order to understand how genes were involved in life processes. This left behind some cardinal problems, such as embryogenesis, i.e. how an undifferentiated cell gives rise to differentiated progeny cells, or how a skin cell transfer the "skin nature" to daughter cells, or the structure of the nervous system whose basic components' behaviour alone cannot explain complex functions, or even some problems pertinent to molecular biology, such as the high number of duplicated genes throughout the genome (especially for higher organisms).

2.3. Reductionism under attack

The first half of the 20th century was marked by the revolution in physics. The violent controversies that accompanied the elaboration of quantum mechanics shed light on uncomfortable aspects of epistemology[2]: following Louis de Broglie, Schrödinger established a vision of matter as waves, whereas primarily the Copenhagen school led by Niels Bohr thought of matter as particles with quantified energy states. The unification of these two theories that were also two visions of Nature occupied all major physicists till the end of the thirties. This unification was done at the cost of huge epistemological compromises or revisions, and produced unexpected results, such as Heisenberg's uncertainty principle, which states that "*the more precisely the position is determined, the less precisely the momentum is known in this instant, and vice versa*". For many scientists and philosophers, the clear consequence of this principle is that the most intimate details of nature cannot be known at any desired precision. At the same time, the unified theory

[2] The concepts about quantum physics exposed in this paragraph may be oversimplified and provocative, but stress is put on their impact (or the impact of how they were perceived) on the history of ideas.

imposed a duality in Nature: objects can be particles or can be waves, and the way we observe them determines their wave or particle nature. This uncomfortable situation is summed up by Einstein himself:

> *"Quantum mechanics is certainly imposing. But an inner voice tells me that this is not yet the real thing. The theory says a lot, but does not bring us any closer to the secrets of the 'Old One'. I, at any rate, am convinced that He is not playing at dice." Letter to Max Born (1926). Albert Einstein.*

Not only did quantum physics pose fundamental problems about the nature of atoms, but dramatically questioned the determinism program that occupied a very strong position in the thirties:

> *"It is not absurd to suppose that psychological laws may eventually be explained in terms of the behaviour of individual neurons of the brain; that the behaviour of individual cells —including neurons— be explained in terms of their biochemical constitution; and that the behaviour of molecules —including the macromolecules that make up living cells— may eventually be explained in terms of atomic physics." Oppenheim and Putnam (1958) [42].*

Cybernetics: influence, misconceptions and promises

Molecular biology did not evolve independently of other disciplines, and was deeply influenced by other branches of research, including information theory: the concepts of information, message, program, are still currently used in molecular biology to explain the respective functions of DNA, RNA and proteins and their interplay (see below for a discussion about the implications of considering DNA as a program). But information theory and cybernetics provided much more to biology than simple vocabulary and metaphors.

Information theory was pioneered by Shannon in the late forties [43]; its major achievement was to provide a mathematical framework for the quantitative evaluation of information, and its modifications through coding and transmission. His concepts were rapidly taken up in the field of biology, primarily to evaluate the information content of whole living beings [44, 45], and a few years later that of DNA [46, 47]. Though relatively unfruitful in practise, these approaches influenced the way molecular biologists perceived DNA as an information-rich message.

During the same period appeared new fields such as computer sciences and cybernetics: computer sciences were developed by people such as von Neumann and Turing during World War II. Not only were they interested in the building of computing machines to break German secret codes, help design nuclear weapons or do weather forecasting through the Pacific Ocean, but they also laid down the principles of theoretical computation. For example, Von Neumann designed the so-called "Von Neumann architecture", in which the computer consists in a set of instructions, a program that will be interpreted by the set of instructions, and data (both input and output of the program). This architecture is still the one used in usual computers. It is worth examining in some details these early fundamental developments in computer sciences, because their influence on molecular biology and genetics is so deep that even in the general audience, DNA is perceived as a computer program for the building of living beings. One of the fundamental aspects of these early developments in computer sciences took over the idea pioneered by Gödel in *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme* for the demonstration of his incompleteness theorem: the basic idea is that formulations about number theory can be themselves formulated as numbers. This trick thus induces a recursivity that is the core of Gödel's proof [48]. Turing and Von Neumann applied this idea to computer architecture: the set of instructions, the program and data share essentially

the same structure and need not be differentiated. Only does the external a priori knowledge of what a set of instructions is, what a program is and what data is (i.e. the hardware structure of the computer) differentiate them.

The popular perception of DNA as a program has grown on these developments of computer sciences as well as on the underlying belief that there need be a master architect to organise living processes. However, no syntax has ever been identified in DNA so far. The only clear established fact is that DNA *codes* for proteins. As pointed out by Henri Atlan [49], if DNA is to be seen as a program, the cellular machinery must be the set of instructions that interprets the program. Algorithm theory specifies that the set of programs interpretable by a given set of instructions cannot be known *a priori*, and that only experiments can tell if a program is interpretable. Interestingly, molecular biology has extensively proven that a given cellular system can "interpret" DNA from various organisms, given proper molecular support is provided. The set of results (structures of the cellular system) is defined by the program; so if each DNA is a program, it will produce its own set of results. Conversely, if DNA is to be considered as data, the cellular machinery must be the program. The set of DNA accepted as input is defined by the program; in practise, the cellular machinery can accept a wide variety of DNA inputs. But this time, the set of results (the cellular structures) is defined by the program that is unique. The set of results will thus be limited to a single class, independently of the input data (the DNA). Remarkably, cellular structures and major biological functions are the same in the whole living kingdom, and the variations depend on the DNA, which corresponds well to the situation in which the program is unique (the cellular machinery) and thus all possible results belong to the same class, the variations depending on the input data. The purpose here is not to show that DNA is a program, or data, but that metaphors from computer

sciences do not describe the reality: the metaphor of DNA as data is as pertinent as that of DNA as a program (and in fact better corresponds to the reality).

The development of computing machines, able to make complex operations, gave rise to the idea that some concepts used in computer science and automate control could be useful to model or interpret biological phenomena; in one of the seminal book about cybernetics [50], Ross Ashby adopts the definition of Wiener, as "the science of control and communication, in the animal and in the machine". It is thus from the very beginning a transversal approach to process control, be it a living process or a man-made process. The main contributions of cybernetics to molecular biology have been summed up by Henri Atlan [51]:

- Life processes can be viewed as information processing channels
- Molecular processes are regulated by feedback loops
- Redundancy provides robustness and information can arise from noise
- Collections of simple units can self-organise into complex systems

As already mentioned, the first point has been relatively unfruitful, except by providing a conceptual framework and a language to molecular biologists. The second point has clearly fostered the models for gene regulation. Figure 5 shows a simplified model of the lactose operator as described by Jacob and Monod [52-55]. This model explains how bacteria express enzymes responsible for lactose metabolism only in the presence of lactose: in the absence of lactose, the repressor gene Lac 1 "produces" a repressor protein that assembles into a tetramer unit. This repressor tetramer unit has a high affinity for the operator DNA portion that is situated just between the promoter site (where RNA polymerase binds) and the genes to be expressed (Lac Z for for β-galactosidase that metabolises lactose, Lac Y for lactose permease, a membrane protein that permits the entry of lactose into the cell, and Lac A for thiogalactoside transacetylase, whose function is not

well documented). In the absence of lactose, the RNA polymerase cannot thus move from the promoter site to the genes themselves and transcription is inhibited. In the presence of lactose, its intracellular form allolactose binds to the repressor tetramer unit that loses its affinity for the operator site. The RNA polymerase can thus move from the promoter site to the genes, and Lac Z, Lac Y and Lac A are transcribed and expressed. Lactose is then consumed, and if not supplied to the cell, its concentration diminishes and it is released from the repressor that can thus bind again to the operator site and inhibit gene transcription.

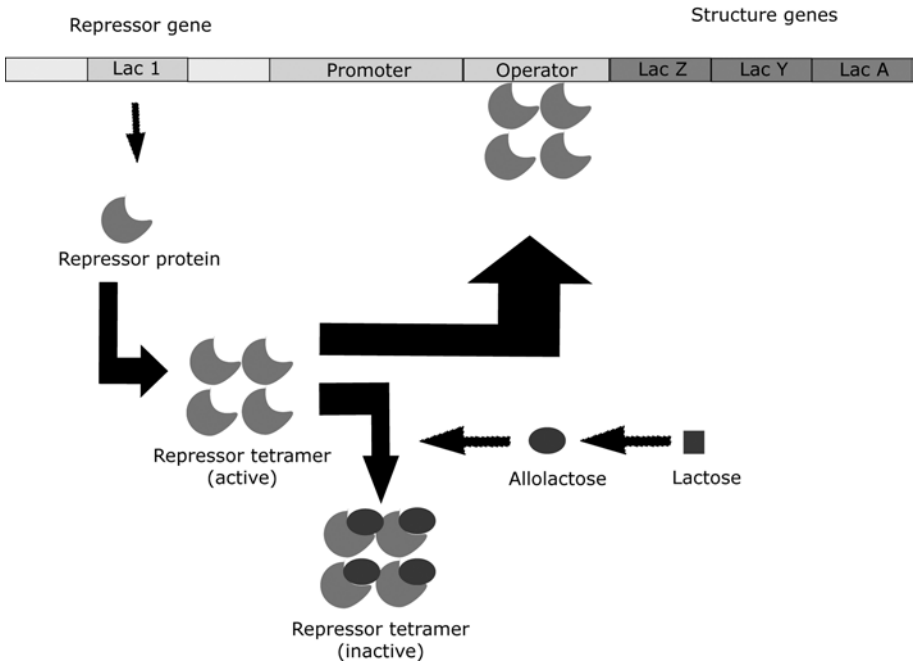


Figure 5. Structure of the lactose operon as described by Jacob and Monod.

(This model is simplified in the sense that it does not explain the metabolism when both glucose and lactose are present. In fact, the RNA polymerase has a relatively low affinity for the lactose promoter site, so that the glucose metabolism machinery is preferentially expressed. When glucose rarefies, the intracellular concentration of cAMP levels up, and cAMP complexes with CAP —the Catabolite Activator Protein—; the

complex CAP-cAMP then binds to the promoter site, which enhances the affinity of the RNA polymerase for the lactose promoter site by a factor of 50.)

This kind of models of feedback regulation where β-galactosidase metabolises lactose, whose disappearance inhibits the production of the enzyme, was already well known in process control and cybernetics [50]. It is hard now to evaluate the cross-fertalisation of cybernetics and molecular biology; gene regulation mechanisms may have been well discovered without any culture in process control, but what is sure is that cybernetics formalism was early applied to molecular biology problems, for example by Griffith [56, 57]. Unfortunately, the utility of such approaches beyond metaphors necessitated reliable time-resolved quantitative data about RNA and protein levels. Only relatively recently were convincing matches between theoretical models and experiments published [58, 59], as shown in Figure 6:

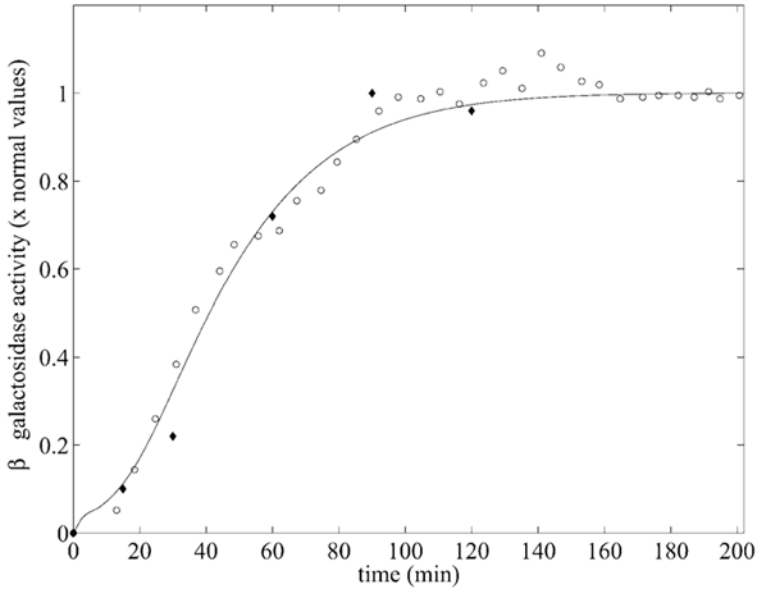


Figure 6. Evolution of β-galactosidase activity in response of lactose supplementation. Open circles and plain diamonds are from two different datasets from the literature. Solid line is the modelled activity. Reprinted with permission from [59].

Cybernetics has also directly contributed to the understanding of biological robustness to external factors through the concept of redundancy. This is probably the domain into which there are the most numerous misconceptions about the influence of cybernetics on biology. In man-made objects and machines, robustness is guaranteed by duplication of key elements (such as triplication of flying commands on an airplane). The fact that living organisms are usually robust to exposure to different growth conditions, stressing environments and mutations has driven many researchers to take for granted that their robustness is due to duplication of essential metabolic pathways or enzymes. This is a very anthropomorphic view of life processes and evolution, which comes from the confusion between species survival and individual survival. As summed up ironically by Nicolas Le Novère:

> *"If I am an oyster, how then can I be sure to preserve my offspring from negative deviations? If my critical subsystems are redundant, I'll be able to survive minor deleterious variations, and I'll pass those variations to my offspring. Following the neutral theory of molecular evolution (Motoo Kimura), these variations can invade the population by genetic drift. Now if my species encounter situation where an optimal function of the subsystem is required, it will be wiped-out from the surface of the earth. Too bad. Another scenario: The redundancy of my critical subsystems has been kept at its minimum. If any of my offspring experience a small deviation in one of those subsystems, it will die quickly. Well, you know what? I don't care. I'm an oyster. I can spawn ten millions eggs. I can afford to loose 90% of them." [60]*

In fact, observation has proven that critical elements are not duplicated in cells; for example, there is a single RNA polymerase to transcribe DNA into mRNA. If there is a mutation in the gene coding for such an essential enzyme, the organism dies before it can

transmit this mutation to an offspring. Wagner has recently investigated possible sources of robustness in living organisms [61]: there are a lot of experimental evidences that key genes (in all species and in all kind of genes, enzymatic, structural or regulatory) can be silenced or deleted without dramatic effect, which for many molecular biologists is a proof that these genes are actually duplicated, or that other gene-products can take over the deleted function. This is again a too reductionist approach to gene function: for example, Zoli *et al* have shown that a mice model lacking the β2 subunit of the nicotinic acetylcholine receptor, which is an essential component of signal transduction, and which was known experimentally to be non essential to proper development, turned out to develop dramatic neurodegeneration during aging [62]. So even asymptomatic knock-out organisms can demonstrate complex physiological behaviours that cannot be explained by single gene deletion. But Wagner investigated another source of robustness, namely the distributed robustness [61]: many parts of a system contribute to its overall function, but all of these parts have different functions. When one part fails or is mutated, the system can compensate for this failure, but not because a back-up redundant part takes over the failed part's role. An example of such a robustness is the role of the enzyme glucose-6-phosphate dehydrogenase (part of the pentose phosphate shunt) in the metabolism of NADPH/NADP in *Escherichia Coli.* In the wild type organism, two thirds of NADPH is produced by the pentose phosphate shunt . Most of NADH is produced by the tricarboxylic acid cycle. If the gene for the glucose-6-phosphate dehydrogenase is mutated or silenced, the oxidative branch of the pentose phosphate shunt is blocked, and thus production of NADPH should be blocked. But an increased flux is observed through the tricarboxylic acid cycle, which generates NADH. This increased NADH is then transformed into NADPH via a massively increased flux through the transhydrogenase reaction. In summary, NADPH can still be produced by recruiting some other pathways of different metabolic networks. This is an

inherent feature of complex distributed systems to be able to compensate for blocked pathways by recruiting different, complementary biochemical networks. This kind of robustness cannot be explained by molecular reductionism alone, and its concept directly comes from cybernetics theory [50].

The example above showed how robustness can appear as an emergent property of a complex system. The concept of emergence[3] has been pioneered in engineering by Alan Turing through the development of neural networks [63]: by mimicking neurons function, Turing proposed new computing machines based on the following elementary units:

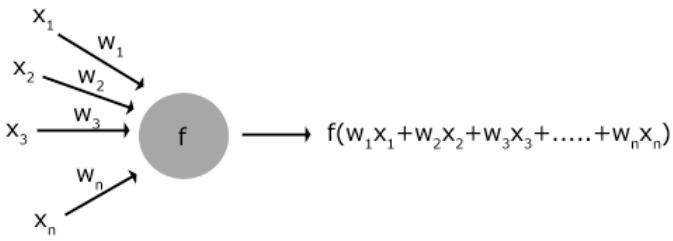


Figure 7. Schematic of an artificial neuron.

Where x_1…..x_n are external inputs or outputs of other neurons, w_1….w_n are the weights of the connections and the output is whether an external output or one input to another neuron. The function f is usually whether a threshold or sigmoid function. When assembled into networks, these artificial neurons can demonstrate relatively complex behaviours, such as pattern recognition abilities (the Perceptron [64]) that are hardly feasible by classical computers. The striking feature of neural networks is that complex responses can be generated by collections of very simple computing units such as the one depicted in Figure 7. As for example, Hopfield *et al* described an elegant computation of the

[3] Emergence is here defined as the appearance of properties that cannot be attributed to parts of the system, but arise from a collective behaviour of these parts.

classical travelling salesman problem (which is a canonical NP – non deterministic polynomial – problem, a class of problems whose solution cannot be computed in a polynomial time): a salesman has to visit n cities while minimising the distances travelled; all the problem is to determine in which order the cities have to be visited [65]. There are still no "classical" algorithm to solve this problem because of the combinatorial complexity of testing all possible pathways, but Hopfield *et al* obtained a solution in a convergence time 10^{23} faster that the combinatorial exploration of all possibilities (if it were possible). With the particular network developed by Hopfield for this problem, a neuron corresponding to the city *i* (row *i*) is activated in column *j* if it has to be visited in j^{th} position in the travel:

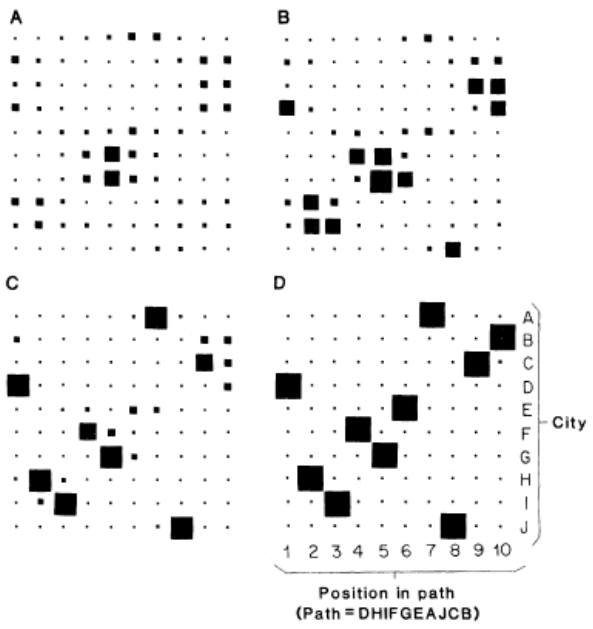


Figure 8. Neural computing of the travelling salesman problem. All dots are artificial neurons. The size of each black box depicts the activation (the output value) of each neuron. Cities are lettered A to J (vertical axis); the order in which they are to be visited are on the horizontal axis. A is the random initial state of the network, B-C intermediary states through computation, and D the final state. Reprinted with permission from [65].

The development of neural networks was a breakthrough in computing science, as well as in biology: in fact the initial ideas for the development of neural networks were taken from biology, and their practical implementation showed biologists how collective properties can emerge from collections of simple units. Interestingly, the concept of emergence had been introduced in the second half of the 19th century by the natural philosopher George Lewes [66], but only in the second half of the 20th century received attention from the biological community [67].

At the beginning of the second half of the 20th century all these approaches at the interface between computer science, process control and biology gave rise to a new discipline: system theory, which was founded by Norbert Wiener, William Ashby and Heinz von Foerster (cybernetics), von Neumann and Alan Turing (connectionist approaches), Jay Forrester (computer sciences), Ludwig von Bertalanffy and Henri Atlan (theoretical biology), Warren McCulloch (neurophysiology), Edgar Morin (sociology)… As popularised by Joël de Rosnay [68], system theory is a holistic approach that consists in looking at interactions between components of a system in order to understand the behaviour of the system as a whole. Whereas its advent may have been relatively unnoticed in the general scientific audience, it consists in a radical paradigmatic change, which can be explained by examining neural networks: a reductionist examination of a neural network would end-up with the description of a single neuron such as the one depicted in Figure 7 together with its activation function, which tells very little on the potential of the whole network, as all the important features are implemented in the connections of the neurons (i.e. in the weights of the connections). Joël de Rosnay further proposed a conceptual instrument called the *Macroscope*: it can be described as a marriage between the microscope and the telescope that allows to look at systems from far away and get a general picture and zoom into the most intimate details of the system. The holistic approach is thus

characterised by zooms in (that allow molecular, reductive description of biological systems) and zooms out (to put all the molecular information into perspective with other levels of organisation).

3. The rise of large-scale biology

3.1. Early history of genome sequencing projects

The discussions about possibilities to sequence whole genomes started in the beginning of the eighties and crystallised in 1985 in a meeting organised by Robert Sinsheimer (chancellor of the University of California at Santa Cruz) featuring John Sulston (Cambridge University) and Robert Waterston (Washington University in Saint Louis) who were already involved in the genome mapping of *C. Elegans*, Bart Barrell (head of large-scale sequencing at the UK Medical Research Council), David Botstein (MIT), Helen Donis-Keller (Collaborative Research Inc.) Walter Gilbert (Nobel Prize in chemistry in 1980 for the development of DNA sequencing methods) and George Church (Harvard University, and Leroy Hood (California Institute of Technology). Whereas the conclusion of the meeting was that large-scale DNA sequencing was not feasible, it triggered Walter Gilbert's interest in taking the lead of assembling a genome sequencing initiative. He soon gained support from Francis Crick and Maurice Wilkins (Nobel Prize in 1953 for the discovery of the DNA double helix) as well as Charles DeLisi, a cancer biologist who was heading the Office of Health and Environmental Research at the Department of Energy (DOE). Several meetings held out in the course of 1986 led to the conclusion that large-scale genome sequencing might be doable, but questions about how desirable it was were at the heart of the building of real projects: scientifically speaking, many "traditional"

molecular biologists were convinced that conventional approaches were more fruitful that brute-force DNA sequencing:

> *"An approach that included mapping, genetics and biochemistry makes a lot more sense". Maxine Singer (National Cancer Institute, Bethesda), 1986. [69]*

In parallel, the scientific task to be accomplished was questioned by many as a non-scientific enterprise; Sydney Brenner (Nobel Prize in Physiology and Medicine 2002 for the elucidation of genetic regulation mechanisms) joked that sequencing is so boring that it should be done by prisoners and not by scientists [69]. Lastly, there was a reasonable concern that a nation- or worldwide effort for genome sequencing would drag all funding for biomedical research and cut out all other projects.

The situation was unblocked by the investments of the US Department of Energy into its own genome sequencing effort at the end of 1986, due to the commitment of Charles DeLisi, the need for DOE to redesign its activities after the decline of military activities, and to its know-how in large scientific projects management. Soon after, a nation-wide genome sequencing effort received political appraisal as a multi-phase project: first establish physical maps of human chromosomes, second evaluate and develop the sequencing technologies on model organisms, and lastly sequence the human genome [70]. The National Institutes of Health (NIH) took the lead over DOE by appointing James Watson as the head of its new Office for Human Genome Research in 1988 [71]. Under the auspices of James Watson, genome sequencing projects clearly took a "basic science" orientation, whose goal was to decipher the Book of Life.

3.2. Speeding up the race for genome sequencing

But in 1991, Craig Venter, who was then heading a large-scale sequencing lab at the National Institute for Neurological Disorders and Stroke came out with the technology of Expressed Sequenced Tags (ESTs), which would enable to focus on the discovery of biomedically relevant genes at high speed but with less precision than other sequencing technologies. This technological strategy together with the aggressive patent filing of Craig Venter created a huge controversy between genome sequencing pioneers that eventually cost his job to James Watson. He was replaced by Francis Collins, an atypical physician by training, who somehow redesigned the objectives of the Human Genome Project to put emphasis on biomedical aspects. But DNA sequencing technologies were still relatively slow, with only incremental progress on the technological side, and Francis Collins chose to focus the efforts on the development of physical and genetic maps that could already be useful to geneticists and physicians. But in 1995 the situation changed radically, when Craig Venter announced that the genome of *Haemophilus Influenzae* had been sequenced in a year by a brand new technique called shotgun sequencing (a technology that has always been rejected by NIH) [72]. From that demonstration, the race between publicly-funded programs (NIH, DOE and MRC) whose aim was to produce a high-quality consensus sequence of the human genome, and Venter's initiative which was more "gene-identification" oriented, raged. The first consequence of that competition was the biomedical claims made by the two project leaders, Francis Collins for the NIH and Craig Venter, in the urge to maintain the money flowing into their own projects. In 1998, the competition received a second kick when Craig Venter announced the creation of Celera Genomics in partnership with Perkin-Elmer who would single-handedly sequence the human genome within three years. This announcement made Francis Collins reshape the NIH Human Genome Project once again: in order to keep the pace imposed by Celera Genomics, the public project would now

produce a lower quality sequence. The goal was two-fold: first to keep in the race imposed by Craig Venter, and second, to publish as much sequence as possible before Celera Genomics could patent it. Finally the incredibly violent rivalry between Craig Venter and Francis Collins found an end in the decision to publish the results of the human genome sequencing efforts the very same day in Nature for the publicly funded project [73] and in Science for Celera Genomics [74]. The beneficial aspect of this competition is clearly the efficiency in genome sequencing efforts: Figure 9 shows the number of genomes sequenced per year.

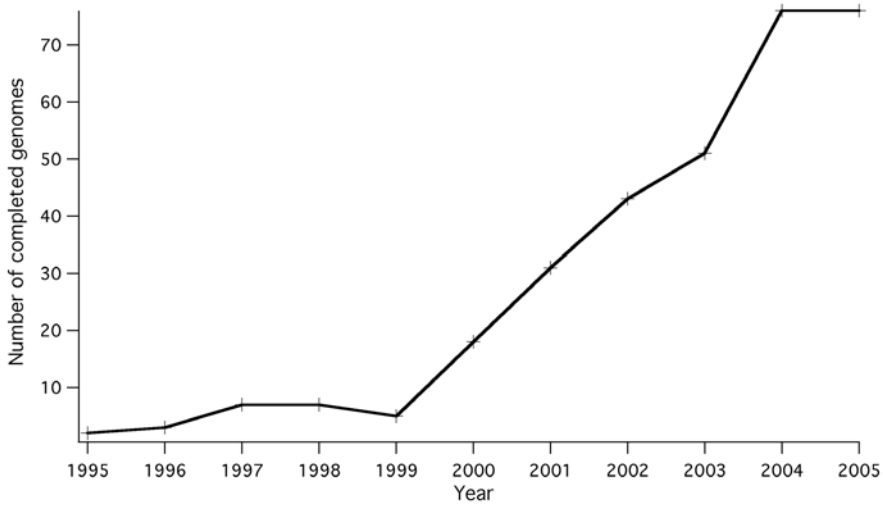


Figure 9. Number of sequenced genomes per year. Information is from http://www.genomesonline.org Data for year 2005 are extrapolated from the first ten months.

There is a clear speed up after 1999 due to technological developments performed during the previous years. In 2005, the number of completed genomes is expected to be the same as in 2004, but there are 760 ongoing sequencing projects for prokaryotes and 531 for eukaryotes.

As summed up by Leroy Hood [75], the Human Genome Project has resulted in four striking features:

- There are approximately 25000-35000 genes in human beings, which is just one third more than in *Caenorhabditis Elegans*, a small worm. How to account for human complexity from this relatively small number of genes is still an unanswered puzzle.
- There are no race-specific genes: the genomes of two Black people may differ more than those of a Caucasian and Black persons. The concept of Race is cultural, not genetic.
- There is a unity throughout the living beings: many fundamental biological systems are similar from bacteria to human beings.
- All living beings are connected: in the human genome, there are 200 or more genes derived from other organisms that cannot be accounted for by evolution (they would have evolved), which means that living beings can incorporate information from surrounding organisms.

These four features were very striking in the sense that they show the unity and similarity of all living beings, contrary to the classical pyramidal view of evolution that places human beings at the top of all living creatures.

3.3. So what?

With the sequence of the human genome (and many other species) being available for more than four years now, claims over its direct biomedical relevance can be re-evaluated: whereas in its first phase, James Watson put the sequencing projects under the auspices of basic research in a philosophical perspective to read the Book of Life, from 1993 Francis Collins put the emphasis on potential biomedical applications. But still today it is hard to find direct biomedical applications of genomics. For example in 2002, Francis Collins pointed out two examples of treatment orientation based on genomic knowledge

[76]: a case of a pregnant woman with predisposition for deep venous thrombosis whose predisposition could be confirmed by genetic testing and thus could be treated with anticoagulants to avoid any problem in the course of pregnancy. The second case cited was that of a four-year old boy treated for lymphoblastic leukaemia with mercaptopurine (a highly toxic compound with potential deadly side effects). Genetic testing revealed that the boy was homozygous for a mutation in the gene that encodes thiopurine S-methyltransferase, an enzyme that inactivates mercaptopurine; drug doses could thus be decreased to levels where no side effects appeared [76]. Francis Collins cited these two examples as canonical of benefits brought about by the genomic revolution. However, this paper rapidly raised criticisms from the medical community, as the two cases would have been treated exactly the same way without any genetic testing [77].

The purpose here is not to claim that genomics has no biomedical value: already demonstrated applications range from the identification of new pathogens (such as the severe acute respiratory syndrome SARS [78-80]), prognosis and therapy guiding in breast cancer [81], stratification of patients with long QT syndrome [82] or myocardial infarction [83]… But genomic testing has proven informative mainly in monogenic diseases, or in diseases in which mutations on one or two genes play a dominant role. However in most diseases or disorders, genetic causes are not so clear or a great number of genes are involved in the disease. For example, whereas most genetic studies on Alzheimer disease have focused on four genes (those coding for presenilin 1 (PS1), presenilin 2 (PS2), amyloid precursor protein (APP) and apolipoprotein E (APOE)) [84], a number of other genes have been reported to play a role in the disease: a simple full text query of *Entrez Gene* from the National Center for Biotechnology Information [85] ("Alzheimer AND Homo Sapiens") retrieves 125 entries; complementarily, a similar search in SwissProt (All Text: Alzheimer and Organism: Homo Sapiens) retrieves 57 different proteins that have

been documented to be involved in Alzheimer disease. Whereas genetic analysis may allow prognosis of the age at onset and progression of the disease [86-88], no breakthrough in disease treatment or curing has been provided yet by genomics; in contrary most studies focus in a reductionist approach to the four genes coding for presenilin 1 and 2, amyloid precursor protein and apolipoprotein E, even when genomic technologies are used.

Besides the reductionist approach that are often at work even when genomic technologies are used, there is a clear limitation of genomics: genes code for proteins, but do not bear any physiological functions by themselves. Proteins embody functions, such as catalysis, structure, transport, signal transduction… This limitation of genomics was early recognised by Charles DeLisi, an early pioneer of genome sequencing at DOE:

> *"But, perhaps, an even more fundamental change has begun. The high-throughput computational and experimental methods of the post HGP* [Human Genome Project] *era are forcing molecular biology away from its spectacularly successful reductionist roots back toward the integrative systems physiology required to understand cell behavior. High-throughput genomic methods, which are a revolution for global characterization, are a start in that direction. The long leap from characterizing to understanding, however, will be possible only after analogous technologies are developed for proteins." Charles DeLisi. Human Genome Program, U.S. Department of Energy, Human Genome News (v11n3-4), July 2001.*

This quotation positions well the situation of genomics: first it opens the way for integrative biology by whole organism study. Secondly, it will achieve its promises only through the advent of proteomics, which is the true level at which functions are embodied.

4. What is proteomics?

Proteomics can now be seen not only as the high-throughput study of proteins or a set of techniques to study proteins, but as a discipline that is complementary and indispensable to genomics. Because many events can occur in the pathway from genes to proteins, one can not necessarily predict that a function (a protein) be present because a gene is expressed. Many different processes regulates protein production even downstream RNA, such as post-translational modifications (PTMs) that can act as a trigger for function (such as is often the case for phosphorylation), truncation, targeting, involvement in multi-protein complexes… As a consequence, genomic or transcriptomic information cannot be taken as a marker of function. Many studies have investigated the correlation between mRNA levels and protein levels, without clear answers:

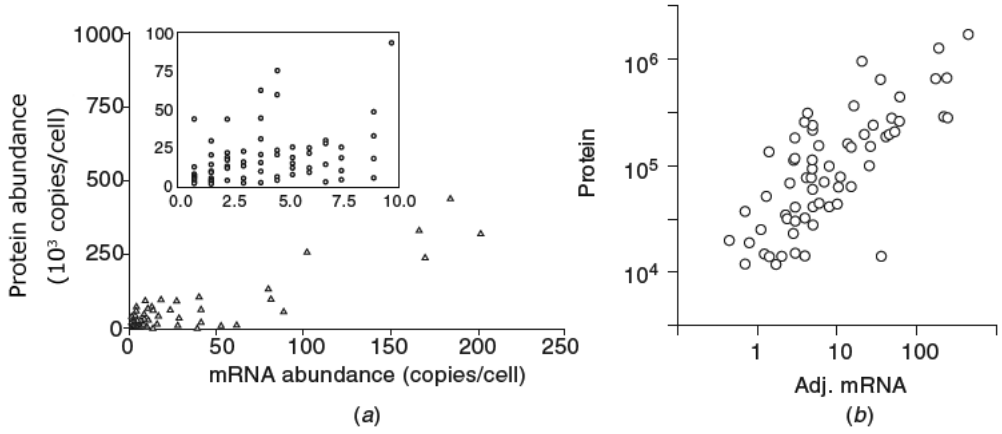


Figure 10. Correlation between protein and mRNA levels in S. Cerevisiae. (a) for 106 genes in yeast. The inset shows the low-end portion of the main figure. The Pearson product moment for the entire data set was 0.935 and 0.356 for the inset. Reprinted with permission from [89]. (b) Correlation of protein abundance with adjusted mRNA levels in yeast. The Pearson product moment is 0.76. Reproduced with permission from [90].

As shown in Figure 10, protein and mRNA levels are often poorly correlated, and relying on transcriptomics to assay functions can be totally misleading due to downstream

regulations of the protein expression process. It is thus essential to characterize organisms at the protein level also, which is again the level of biological processes.

In paragraph 1, proteomics was defined as "the identification of all proteins expressed by a genome, cell or tissue, their quantification, and determination of their interactions". After the detour of paragraphs 2.2 and 2.3 about paradigmatic changes imported from cybernetics and systems theory into biology, the definition can be rephrased in terms of systems biology: proteomics is the study of biological systems at the level of proteins. It consists only in a single-level description of a system, and as such cannot claim to be *the* necessary description of a living being. It is not more pertinent *per se* than genomic, metabolic, anatomical, physiological… descriptions. Above all it is not the cataloguing of proteins present in a given sample. It is one part of systems or integrative biology, which consists in putting in perspective all the molecular information collected at the DNA, RNA or protein level with more physiological information to answer a particular question. Whereas this definition may read trivial and sound like "systems biology is the science of everything", it truly consists in a radical change in the way biological research proceeds.

One difficulty of this new approach is to choose the pertinent level of description to investigate a particular problem. Is Alzheimer disease to be explained at the gene level? This is still an unanswered question: some properly designed genomic studies may unravel complex genomic predispositions that account for a large proportion of disease appearances. Proteomic studies may elucidate that the disease appear as a consequence of the accumulation of particular proteins in a particular state that have nothing to do with genes but depend only on altered protein processing mechanisms. Environmental studies may prove that there are external factors that are essential for the onset of the disease. Further studies may conclude that there are genetic, epigenetic and environmental factors

that must act together for the disease to appear; in this case, only system approaches have a chance to explain the disease. It is highly probable (or at least the conviction of many) that many systemic and multifactorial diseases such as diabetes, cardiovascular diseases, cancer… will be explained only when investigated by systemic approaches: one of the most convincing biomedical progress in genomics is the predisposition for breast cancer: if a woman has a given mutation in the gene BRCA1, she has a probability of 70% of developing a breast cancer before the age of 60. Even with such a monogenic trait, the risk of getting cancer is not 100%, because there are other factors at work that prevent or promote the appearance of a breast cancer.

The choice of the right level of description of an organism in perspective of a particular problem is a difficult task. In April 2004 was held in London the Second conference of the Centre for Economic and Social Aspects of Genomics; a large number of talks was devoted to the field of nutrigenomics, whose interest is to develop new foods or new diets to improve general health based on genomic information. Beyond all biological aspects, all studies were somehow inconclusive because they missed a more social dimension: what is health? Is health the absence of disease? Is health a feeling of wellness? The concept of health is highly cultural and very personal [91], and any strategy to improve health must comprise this social aspect in order to properly target the desired effect. As such, integrative biology is not limited to biological sciences, but may also incorporate social sciences approaches in particular cases.

As outlined by Leroy Hood [75] and Francis Collins [92], the biomedical promises of integrative biology lie in predictive and preventive medicine. It is already possible to test for disease predisposition, such as genetic tests for BRCA1 mutations. It may soon be possible to test for many disease predispositions through genetic testing or more conventional protein-based assays. But prediction without cure is an anathema to medicine.

"Cures" can be envisioned of two different classes: preventive medicine may be able to perform individual counselling to limit exposure to environmental of behavioural factors that may contribute to the onset of a predisposed disease. At the other end, there is a strong hope that integrative biology will allow early diagnosis of disease onset, and the design of more efficient treatments. The rise of proteomics has (re)opened the quest for disease biomarkers. Most studies focus on the detection of biomarkers in easily accessible biological fluids such as plasma or urine.

Whatever the technology used, the strategy is always the same: samples from patients with the disease under investigation are compared with samples from control individuals (both healthy individuals and patients with related diseases). Molecules (DNA in genomic studies, mRNA in transcriptomic studies or proteins in proteomic studies) are measured and quantified in the search for quantitative differences between disease and control samples. When a statistical difference is found, the corresponding molecules become disease biomarkers than can potentially be assessed in routine clinical labs for diagnosis. In classical reductionist approaches, molecules under investigations are usually subclasses of biomolecules that have an a priori relevance to the disease. In systems biology, all molecules are investigated; this poses a fundamental problem, as no (or very little) a priori knowledge of the disease is incorporated in the study to keep a chance to identify radically new biomarkers. The drawback of the approach lies in its blindness: the identified biomarkers may or may not have a clear deterministic link with the disease. A now classical example is the discovery of ovarian cancer biomarkers by Petricoin *et al*: the authors used a recently developed analytical strategy called surface-enhanced-laser-desorption-ionisation (SELDI) mass spectrometry to get protein profiles of plasma from women diagnosed with different types of ovarian cancers and compare them from serum

samples of healthy women. The experimental outputs of the study are quantitative protein profiles such as the ones presented in Figure 11.

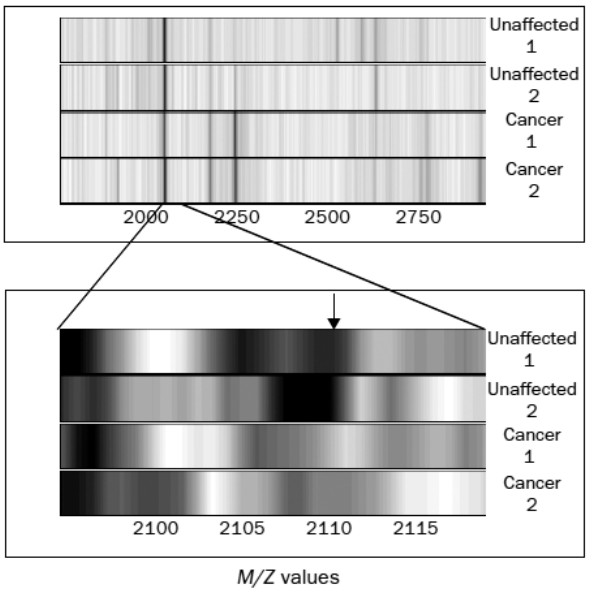


Figure 11. SELDI-TOF-MS protein profile of controls (Unaffected 1 and 2) and patients with ovarian cancer (Cancer 1 and 2). Peptide and protein m/z ratios are on the horizontal axis; intensity is proportional to the normalised protein abundance. Reprinted with permission from [93].

Computer algorithms are then used to identify statistically significant differences in protein abundances between unaffected and cancer patients. Five peaks at *m/z* values of 534, 989, 2111, 2251 and 2465 were sufficient to discriminate between disease and control samples. In a second step, this set of five biomarkers was measured in other samples in a blind study; the diagnostic power of the test is shown below:

	Classification by proteomic pattern	
	Cancer	Unaffected
Unaffected women		
No evidence of ovarian cysts	2/24	22/24
Benign ovarian cysts < 2-5 cm	1/19	18/19
Benign ovarian cysts >2-5 cm	0/6	6/6
Benign gynaecological disease	0/10	1/10
Non gynaecological inflammatory disorder	0/7	0/7
Women with ovarian cancer		
Stage I	18/18	0/18
Stage II, III, IV	32/32	0/32

Table 1. Classification of serum samples by SELDI-TOF-MS. From [93].

As shown in Table 1, the test classifies correctly all patients with ovarian cancer (100% sensitivity over 50 patients, with correct segmentation between Stages I and II-IV), as well as non-cancer patients (95% selectivity with further good segmentation between different gynaecological disorders). In total, the positive predictive value is 94%; in comparison, the state-of-the-art immunoassay for ovarian cancer performed on this second set of samples yielded a positive predictive value of 35%, which means that this test generates a lot of false positives. These results of course appeared as a breakthrough in diagnostics, in particular in the correct detection of stage I cancers, and a magnificent demonstration of the power of proteomic approaches. This work won immediate acclaim: in addition to the congressional resolution urging further funding for this research, the consumer magazine Health named the test one of the ten top medical advances of year 2002. However, a somehow unsatisfactory point is that the identified biomarkers are indeed not identified: only their *m*/*z* ratio is known, but their identity and functions are unknown.

This blind approach turned out to be full of pitfalls: the paper by Petricoin *et al* [94] received early critics: several authors contested the statistical analysis of data and biomedical conclusions raised from it in the context of general screening programs [95-97];

more importantly, Diamandis raised fundamental questions about the methodology used by Petricoin *et al*: "classical" cancer biomarkers (such as prostate-specific antigen (PSA), α-fetoprotein, carcinoembryonic antigen) are known to be present in blood at concentration in the range of 1-10 ng/mL (corresponding to 35-350 pM for PSA, 15-150 pM for α-fetoprotein, and 0.2-2 nM for carcinoembryonic antigen), which are usually way below the sensitivity of mass spectrometers. Even when solid-phase fractionation is used as is the case in SELDI, it is highly doubtful that this low concentration tumour-derived proteins can be isolated from high-abundant proteins and concentrated enough to be detectable by mass spectrometry. Diamandis thus hypothesised that the biomarkers used by Petricoin *et al* are epiphenomena of metabolic changes due to the presence of the tumour (such as inflammatory proteins expressed by liver as a systemic response to the tumour) [98]. Unfortunately the results presented by Petricoin in his two seminal papers about ovarian [93] and prostate [99] cancers turned out to be flawed by methodological and experimental biases [100-104]. The situation was further complicated by patenting and marketing issues of the test for ovarian cancer (OvaCheck™) [105]. Despite the controversy about Petricoin's work, there are now tens of publications about proteomic diagnosis of various types of cancers, mainly focused on analytical developments; unfortunately there are no large-scale clinical study to validate potential biomarkers, and new biomarkers discovered by this approach are rarely properly identified and characterised. A noticeable exception is the work of Koomen *et al* [106]: they used solid-phase capture of pancreatic cancer and control plasma samples with sequential elution to fractionate proteomic samples (not the SELDI approach) before analysis by MALDI-MS. After selection of peaks that allowed discrimination between disease and control samples, corresponding biomarkers were identified by subsequent proteomic characterisation. The six biomarkers that allowed classification of samples were identified as α-1-antitrypsin, apolipoprotein A-I, hemoglobin

β, inter-α-trypsin inhibitor, serum amyloid A, and haptoglobulin. Among these, serum amyloid A and haptoglobulin relates to acute phase response of the host, but do not have any connection with pancreatic cancer. Conversely, plasma protease inhibitors have been documented to be linked to pancreatic cancer, but are also associated with acute phase response, which cast doubts about their specificity as diagnostic markers of pancreatic cancer. These results revived the controversy about the potential of proteomics for biomarker discovery: Diamandis and van der Merwe explained their doubts about the potential of the methodology by pointing out that each identified biomarker shows small differences in abundance between healthy and disease samples compared to classical cancer biomarkers used in diagnostics [107]. At the same time, Petricoin's position is that:

> *"Perhaps the best biomarkers for early detection* [of cancer] *will not even be tumor-derived molecules*. […] *Diamandis claims that proteomic patterns should be based on proteins of known identity. Our data show, to the contrary, that the diagnostic usefulness of proteomic pattern technology is independent of the underlying identity of the proteins. This finding may constitute a new paradigm, in which proteomics coupled to bioinformatics could have direct clinical impact in the clinic."[94]*

The fact that Petricoin's data turned out to be flawed complicates the settlement of the controversy. On one part, the controversy lies in the opposition between reductionist and holistic approaches that are exemplified in the differences by proteomic-based diagnostic approaches and classical immunoassays. It is unlikely that the case be settled rapidly; the validation of proteomic-based diagnostic procedures will most probably necessitate large-scale studies where not only disease and healthy samples will be compared, but other disease samples will be also included: as proteomic-based biomarkers seem to be systemic

responses to disease in most cases, they will have to be specific for the disease under investigation. This means that their specificity will have to be tested on samples from very different diseases producing general systemic responses.

5. Can proteomics identify new biomarkers?

As discussed above, the quest for new biomarkers has been boosted by the advent of genomics and proteomics. Large research programs have been launched in the diagnostics and pharmaceutical industry (whose pipeline is running short of new viable drugs [108]) to identify diagnostic biomarkers, therapy targets, and biomarkers for drug validation and treatment follow-up. The major driving force for such programs is the ability to look at all kinds of genes (in a genome) and all kinds of proteins (in a proteome) at the same time, and thus to potentially discover unexpected biomarkers, or identify sets of biomarkers that can serve together for patient discrimination, and treatment follow-up. Unfortunately, even in the simplest case — the search for diagnostic biomarkers—, these blind approaches have proven to be full of pitfalls (like the ovarian cancer diagnosis by plasma profiling presented above), and it is now more or less consensual that biomarkers need be identified and their physiological function need be documented in order to provide some security in diagnostics [109], even to be used in diagnostics. The case is even more complex when biomarkers become potential therapeutic targets: the background concept is that if a molecule is present in significantly higher or lower amounts in a disease samples compared to healthy samples, the molecule probably plays a role in the disease. The weakness of this concept has been early recognised in the drug discovery process; Figure 12 lists cases when a biomarker fails to be used as a disease target.

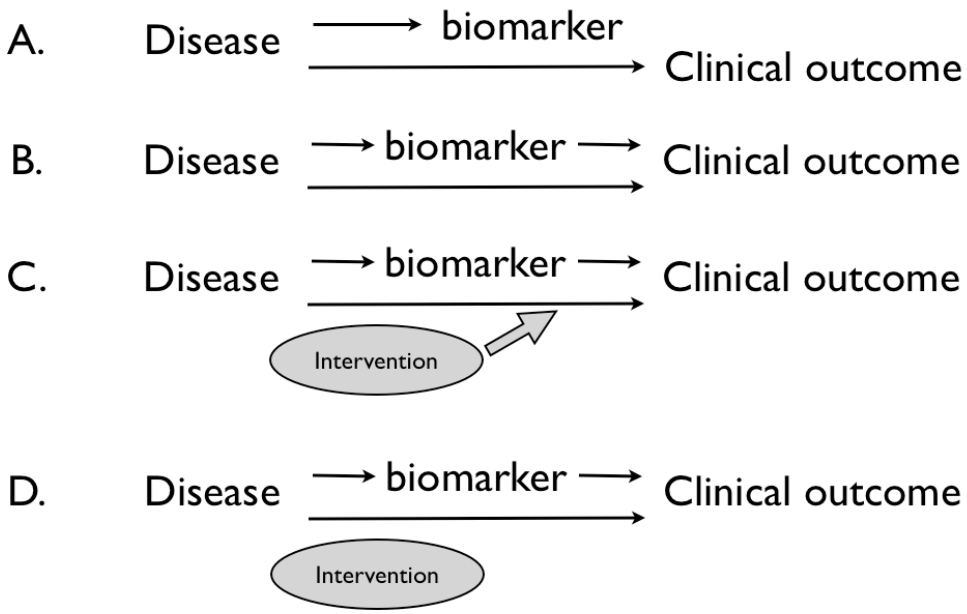


Figure 12. Causes of failure of biomarkers. A. The biomarker is not in the causal pathway of the disease. B. The biomarker is on one of the causal pathways of the disease. C. The intervention (drug) affects one pathway to which the biomarker does not belong to: the treatment cannot be followed. D. The intervention does not have any effect on any causal pathway. Adapted from [110].

When –omics technologies are used only because of their wide capacity to analyse all molecules present in a sample (be it DNA, RNA, proteins or metabolites), they do not bring anything new in terms of biomarker identification, target validation and drug development process. All potential biomarker may eventually fall into one of the categories depicted in Figure 12, and the only benefit of –omics technologies is to provide much more candidates to be tested with the very same approaches that have been used in drug discovery in the last decades; as the limiting step is the drug discovery process lays in clinical trials, identifying more potential biomarkers and targets without significant additional choice-

making information may just overflow drug pipelines. This limitation has been identified as the major cause for drug development failure in the recent years [111].

On the contrary, there is a strong conviction that basic research empowered by –omics technologies may eventually lead to identification of more biomarkers and treatment targets, but more importantly provide more rationale for decision making in order to select the most promising targets based on sound biological knowledge early in the drug development process. More precisely, one of the promise of systems biology is the description and identification of all physiological pathways; when this information will be available, researchers will be able to screen putative biomarkers and drug targets in terms of their role in the disease mechanism: biomarkers that are consequences of the disease but not directly involved in pathways central to the disease (Figure 12 A.) may be used in diagnostics but would be *a priori* rejected as potential drug targets. The identification of parallel disease pathways (Figure 12 B. and C.) would lead to better drug targeting or better treatment efficacy assessment. Again, as was the case a few years ago at the rise of genomics, decision makers have to rely on technological and scientific promises of systems biology paradigm.

6. Conclusion

In this chapter, a very brief and partial history of life science research was provided. Its purpose was to show that whereas reductionism has been the mainstream conceptual framework for scientific research for more than three centuries, other approaches have been introduced through ages in medicine or life science research: Paracelsus' view of disease (disease is a disequilibrium in systemic chemical processes), Claude Bernard's theory of inner medium, Darwin's theory of evolution, are good examples of reactions to reductionism and more holistic approaches to biological processes. The recent advent of

large-scale systems biology is a new example of paradigmatic change; it is definitely not the first one in the balance between reductionism and holism. Its revolutionary aspect lies in the development of new tools and methodologies to acquire large-scale molecular data at the DNA, RNA and protein level.

At the same time, systems biology did not appear *ex nihilo*: it emerged in a context when several epistemological attacks had weakened the reductionist dogma: quantum physics had dramatically changed the vision of Nature as composed of simple elementary units. Cybernetics and systems theory have introduced a science of complexity, through concepts of regulation of elaborated systems, emergence of complex behaviours from connectionist systems, or modelling of biological responses. But systems biology could not have emerged without the success of genome sequencing programs, that showed that large-scale biology was feasible and provided the basic information that was necessary to empower transcriptomics and proteomics.

However, the biomedical outputs of genomics and proteomics have been relatively limited to date; as genes do not actually embody functions, they may be only indicators for a predisposition for disease. On the other hand, proteomics has the potential to detect disease at the true level where molecular processes occur. Unfortunately, proteomic technologies are not mature enough to evaluate the real potential of proteomic diagnostics: though the controversy is not completely settled, it seems relatively clear that tumour-derived proteins are detectable in blood at very low concentration (fM to pM) that are out of reach of current proteomic technologies, at least in routine without massive preconcentration strategies. Conversely, higher abundance proteins that are systemic responses to the disease may serve as biomarkers; but establishing their specificity as diagnostic biomarkers will require much more work than for testing the specificity of a tumour-derived molecule, because these biomarkers will have to be validated against many

other diseases that the one for which they may serve as biomarkers. At the same time, in order for a tissue-specific alteration (such as prostate or ovarian cancer) to induce a systemic host response will probably preclude early stage diagnosis. Efforts in analytical technologies will probably take many more years before proteomics can reveal its potential as a diagnostic and drug discovery tool.

With the relatively limited success of genomic technologies to produce biomedical outputs to date, rises the question about the role of analytical chemistry in the new systems biology. Is systems biology and all its analytical developments just the new hype in basic research, or is there a real promise of systems biology in terms of general health? The answer to this question is still to be determined, and it is probably the role of academic institutions to explore the potential of these technologies. What is sure is that systems biology will rely on molecular information, the quality of which will be crucial to further interpretation and modelling. The example of ovarian cancer diagnosis by plasma proteome profiling shows how crucial experimental design, properly operated instrumentation and critical data analysis are to bring biomedical relevance to proteomic data. At the same time, the discussion about biomarker research (paragraph 5) shows that large-scale analytical technologies do not bring much to biomarker discovery *per se*, but that more fundamental knowledge, interpretation and modelling will be necessary to exploit the true potential of systems biology's paradigm.

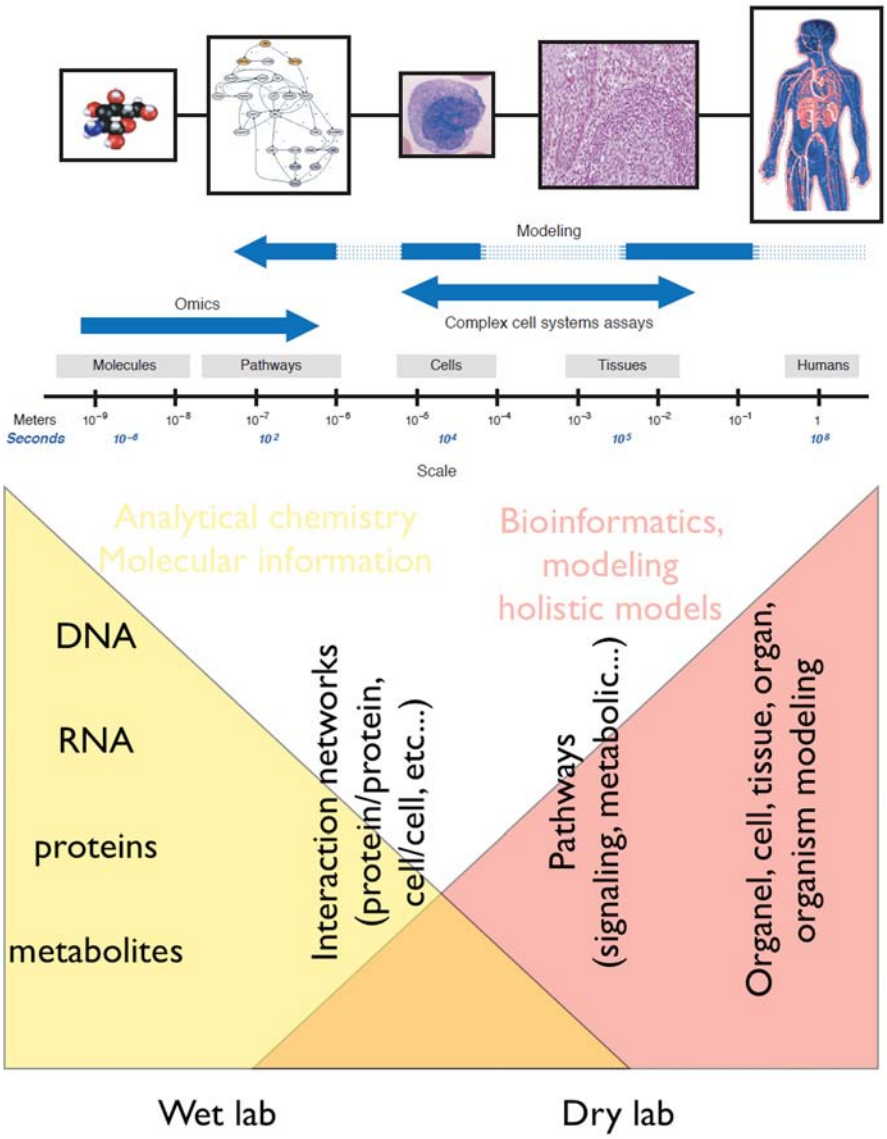


Figure 13. Collaboration between analytical chemistry/omic technologies and modelling. Top is reprinted from [111] with permission.

Figure 13 illustrates the whole disciplines range involved in systems biology: on the left, information is molecular; in this domain, analytical chemistry plays a major role in designing experiments, acquiring data, curing and analysing data to produce, for example, a list of differentially translated proteins. While integrating data into more holistic information such as interaction networks, biological pathways… bioinformatics and modelling play a more and more important role whereas molecular information becomes

less visible. Only by the alliance between the wet and dry labs will systems biology be able to provide sound information about putative diagnostic biomarkers and drug targets.

The goal of this first chapter was to give a brief and partial overview of the concepts at work in proteomics. This thesis work concerns a very focused aspect of one possible analytical strategy used in proteomics, namely the hyphenation of microfluidic devices with electrospray mass spectrometry. As such, it could be integrated in the traditional shotgun approach depicted in Figure 2 to hyphen liquid-phase separations with ESI-MS/MS. As discussed in the next chapter, the main beneficial aspect of using microfluidic systems in place of classical electrospray sources is the potential reproducibility and disposability provided by microtechnological processing. Chapter 3 and 4 provides experimental evidence that polymer microfluidic electrospray sources can be used in various aspects of proteomic research. But microtechnological processing also allows to produce electrospray sources with additional analytical functions, such as on-line sample desalting (Chapter 5) and integration of a tunable sheath flow (Chapter 6). Finally, bioinformatic simulations demonstrate how two analytical technologies developed in our laboratory, namely Off-Gel electrophoresis for isoelectric separation of peptides, and on-line electrochemical tagging of cysteine residues during electrospray ionisation from a microfluidic electrospray source, can be integrated to provide a whole proteomic analytical platform for higher throughput proteome profiling (Chapter 7).

7. References

[1] Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules, *Science*, **1989**, 246(4926): 64-71.

[2] Hillenkamp, F., Unsold, E., Kaufmann, R., Nitsche, R., Laser microprobe mass analysis of organic materials, *Nature*, **1975**, 256(5513): 119-120.

[3] Karas, M., Bahr, U., Laser desorption mass spectrometry, *Trac-Trends Anal. Chem.*, **1986**, 5(4): 90-93.

[4] Nakanishi, T., Okamoto, N., Tanaka, K., Shimizu, A., Laser-desorption time-of-flight mass spectrometric analysis of transferrin precipitated with antiserum - a unique simple method to identify molecular weight variants, *Biol. Mass Spectrom.*, **1994**, 23(4): 230-233.

[5] Fenn, J. B., Electrospray wings for molecular elephants (Nobel lecture), *Angew. Chem.-Int. Edit.*, **2003**, 42(33): 3871-3894.

[6] Tanaka, K., The origin of macromolecule ionization by laser irradiation (Nobel lecture), *Angew. Chem.-Int. Edit.*, **2003**, 42(33): 3860-3870.

[7] Ramsey, R. S., Ramsey, J. M., Generating electrospray from microchip devices using electroosmotic pumping, *Anal. Chem.*, **1997**, 69(6): 1174-1178.

[8] Xue, Q. F., Dunayevskiy, Y. M., Foret, F., Karger, B. L., Integrated multichannel microchip electrospray ionization mass spectrometry: Analysis of peptides from on-chip tryptic digestion of melittin, *Rapid Commun. Mass Spectrom.*, **1997**, 11(12): 1253-1256.

[9] Xue, Q., Foret, F., Dunayevskiy, Y. M., Zavracky, P. M., McGruer, N. E., Karger, B. L., Multichannel microchip electrospray mass spectrometry, **1997**, 69(3): 426-30..

[10] Lion, N., Rohner, T. C., Dayon, L., Arnaud, I. L., Damoc, E., Youhnovski, N., Wu, Z. Y., Roussel, C., Josserand, J., Jensen, H., Rossier, J. S., Przybylski, M., Girault, H. H., Microfluidic systems in proteomics, *Electrophoresis*, **2003**, 24(21): 3533-3562.

[11] Lion, N., Reymond, F., Girault, H. H., Rossier, J. S., Why the move to microfluidics for protein analysis?, *Curr. Opin. Biotech.*, **2004**, 15(1): 31-37.

[12] Rossier, J. S., Youhnovski, N., Lion, N., Damoc, E., Reymond, F., Girault, H. H., Przybylski, M., Thin-chip microspray system for coupling with high performance Fourier-Transform Ion Cyclotron Resonance mass spectrometry, *Ang. Chem. Intl. Ed.*, **2003**, 42(1): 53-58.

[13] Rohner, T. C., Lion, N., Girault, H. H., Electrochemical and theoretical aspects of electrospray ionization, *Phys. Chem. Chem. Phys.*, **2004**, 6(12): 3056-3068.

[14] Lion, N., Gellon, J. O., Girault, H. H., Flow rate characterization of microfabricated microspray emitters, *Rapid Comm. Mass Spectrom.*, **2004**, 18(14): 1614-1620.

[15] Bindila, L., Froesch, M., Lion, N., Vukelic, Z., Rossier, J. S., Girault, H. H., Peter-Katalinic, J., Zamfir, A., A thin chip microsprayer system coupled to Fourier transform ion cyclotron resonance mass spectrometry for glycoconjugate screening, *Rapid Comm. Mas Spectrom.*, **2004**, 18(23): 2913-2920.

[16] Zamfir, A. D., Lion, N., Vukelic, Z., Bindila, L., Rossier, J., Girault, H. H., Peter-Katalinic, J., Thin chip microsprayer system coupled to quadrupole time-of-flight mass spectrometer for glycoconjugate analysis, *Lab Chip*, **2005**, 5(3): 298-307.

[17] Zamfir, A., Bindila, L., Lion, N., Allen, M., Girault, H. H., Peter-Katalinic, J., Chip electrospray mass spectrometry for carbohydrate analysis, *Electrophoresis*, **2005**, 26(3650-3673.

[18] Lion, N., Gobry, V., Jensen, H., Rossier, J., Girault, H. H., Integration of a membrane-based desalting in a microfabricated disposable polymer injector for mass spectrometric protein analysis, *Electrophoresis*, **2002**, 23(20): 3583-3588.

[19] Lion, N., Gellon, J. O., Jensen, H., Girault, H. H., On-chip protein sample desalting and preparation for direct-coupling with electrospray ionization mass spectrometry, *J. Chromatogr. A*, **2003**, 1003(1-2): 11-19.

[20] Henzel, W. J., Watanabe, C., Stults, J. T., Protein identification: The origins of peptide mass fingerprinting, *J. Am. Soc. Mass Spectrom.*, **2003**, 14(9): 931-942.

[21] Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Boucherie, H., Mann, M., Linking genome and proteome by mass spectrometry: Large-scale identification of yeast proteins from two dimensional gels, *Proc. Natl. Acad. Sci. U. S. A.*, **1996**, 93(25): 14440-14445.
[22] Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y., Aebersold, R., Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology, *Proc. Natl. Acad. Sci. U. S. A.*, **2000**, 97(17): 9390-9395.
[23] Rabilloud, T., Two-dimensional gel electrophoresis in proteomics: Old, old fashioned, but it still climbs up the mountains, *Proteomics*, **2002**, 2(1): 3-10.
[24] Wolters, D. A., Washburn, M. P., Yates, J. R., An automated multidimensional protein identification technology for shotgun proteomics, *Anal. Chem.*, **2001**, 73(23): 5683-5690.
[25] Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., Aebersold, R., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags, *Nat. Biotechnol.*, **1999**, 17(10): 994-999.
[26] Sadygov, R. G., Cociorva, D., Yates, J. R., Large-scate database searching using tandem mass spectra: Looking up the answer in the back of the book, *Nat. Methods*, **2004**, 1(3): 195-202.
[27] Douglas, D. J., Frank, A. J., Mao, D. M., Linear ion traps in mass spectrometry, *Mass Spectrom. Rev.*, **2005**, 24(1): 1-29.
[28] Pieper, R., Gatlin, C. L., Makusky, A. J., Russo, P. S., Schatz, C. R., Miller, S. S., Su, Q., McGrath, A. M., Estock, M. A., Parmar, P. P., Zhao, M., Huang, S. T., Zhou, J., Wang, F., Esquer-Blasco, R., *et al.*, The human serum proteome: Display of nearly 3700 chromatographically separated protein spots on two-dimensional electrophoresis gels and identification of 325 distinct proteins, *Proteomics*, **2003**, 3(7): 1345-1364.
[29] Kuhn, T., The structure of scientific revolutions, Chicago University Press, 1962.
[30] Strange, K., The end of "naive reductionism": rise of systems biology or renaissance of physiology?, *Am. J. Physiol.-Cell Physiol.*, **2005**, 288(5): C968-C974.
[31] Diamandis, E. P., Clinical applications of tumor suppressor genes and oncogenes in cancer, *Clin. Chim. Acta*, **1997**, 157-180.
[32] Scheuer, L., Kauff, N., Robson, M., Kelly, B., Barakat, R., Satagopan, J., Ellis, N., Hensley, M., Boyd, J., Borgen, P., Norton, L., Offit, K., Outcome of preventive surgery and screening for breast and ovarian cancer in BRCA mutation carriers, *J. Clin. Oncol.*, **2002**, 1260-1268.
[33] Dillon, D. A., Howe, C. L., Costa, J., Bosari, S., The molecular biology of breast cancer: Accelerating clinical applications, *Crit. Rev. Oncogen.*, **1998**, 125-140.
[34] Descartes, R., Discourse on method, 1637.
[35] Hacking, I., Do we see through a microscope, *Pac. Philos. Q.*, **1981**, 62(4): 305-322.
[36] Pichot, A., Histoire de la notion de gène, Flammarion, 1999.
[37] McCarty, M., Avery, O. T., Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types .2. Effect of Desoxyribonuclease on the Biological Activity of the Transforming Substance, *J. Exp. Med.*, **1946**, 83(2): 89-96.
[38] Pauling, L., Corey, R. B., The Structure of Synthetic Polypeptides, *Proc. Natl. Acad. Sci. U. S. A.*, **1951**, 37(5): 241-250.
[39] Pauling, L., Corey, R. B., Branson, H. R., The structure of proteins - 2 Hydrogen-bonded helical configurations of the polypeptide chain, *Proc. Natl. Acad. Sci. U. S. A.*, **1951**, 37(4): 205-211.

[40] Pauling, L., Corey, R. B., Configuration of polypeptide chains, *Nature*, **1951**, 168(4274): 550-551.
[41] Watson, J. D., Crick, F. H. C., Molecular structure of nucleic acids - a structure for deoxyribose nucleic acid, *Nature*, **1953**, 171(4356): 737-738.
[42] Oppenheim, P., Putnam, H., Unity of science as a working hypothesis. In *Concepts, theories, and the mind-body problem*. Ed. Feigl, H., Scriven, M., Maxwell, G. 1958. University of Minnesota Press, Minneapolis, pp 3-36.
[43] Shannon, C. E., A mathematical theory of communication, *Bell System Technical Journal*, **1948**, 27(379-423.
[44] Linschitz, H., The information content of a bacterial cell. In *Information theory in biology*. Ed. Quastler, H. 1953. University of Illinois Press, Urbana, Ill., pp 251-262.
[45] Dancoff, S. M., Quastler, H., The information content and error rate of living things. In *Information theory in biology*. Ed. Quastler, H. 1953. University of Illinois Press, Urbana, Ill., pp 263-273.
[46] Gatlin, L. L., Information content of DNA, *J. Theor. Biol.*, **1966**, 10(2): 281-&.
[47] Gatlin, L. L., Information content of DNA .2., *J. Theor. Biol.*, **1968**, 18(2): 181-&.
[48] Hofstadter, D. R., Gödel, Escher, Bach: an eternal golden braid, Basic Books, 1979.
[49] Atlan, H., La fin du "tout génétique"?, INRA ed., Paris, 1999.
[50] Ashby, W. R., An introduction to cybernetics, Chapman & Hall Ltd, London, 1956.
[51] Atlan, H., L'organisation biologique et la théorie de l'information, Hermann, Paris, 1972.
[52] Pardee, A. B., Jacob, F., Monod, J., Genetic control and cytoplasmic expression of inducibility in the synthesis of beta-galactosidase by E-Coli, *J. Mol. Biol.*, **1959**, 1(2): 165-178.
[53] Monod, J., Changeux, J. P., Jacob, F., Allosteric proteins and cellular control systems, *J. Mol. Biol.*, **1963**, 6(4): 306-&.
[54] Jacob, F., Monod, J., Genetic regulatory mechanisms in synthesis of proteins, *J. Mol. Biol.*, **1961**, 3(3): 318-&.
[55] Jacob, F., Monod, J., On regulation of gene activity, *Cold Spring Harbor Symp. Quant. Biol.*, **1961**, 26(193-&.
[56] Griffith, J. S., Mathematics of cellular control processes .I. Negative feedback to one gene, *J. Theor. Biol.*, **1968**, 20(2): 202-&.
[57] Griffith, J. S., Mathematics of cellular control processes .2. Positive feedback to one gene, *J. Theor. Biol.*, **1968**, 20(2): 209-&.
[58] Santillan, M., Mackey, M. C., Influence of catabolite repression and inducer exclusion on the bistable behavior of the lac operon, *Biophys. J.*, **2004**, 86(3): 1282-1292.
[59] Yildirim, N., Mackey, M. C., Feedback regulation in the lactose operon: A mathematical modeling study and comparison with experimental data, *Biophys. J.*, **2003**, 84(5): 2841-2851.
[60] Le Novère, N., http://www.ebi.ac.uk/~lenov/index.html.
[61] Wagner, A., Distributed robustness versus redundancy as causes of mutational robustness, *Bioessays*, **2005**, 27(2): 176-188.
[62] Zoli, M., Picciotto, M. R., Ferrari, R., Cocchi, D., Changeux, J. P., Increased neurodegeneration during ageing in mice lacking high-affinity nicotine receptors, *Embo J.*, **1999**, 18(5): 1235-1244.
[63] Teuscher, C., Turing's connectionism: an investigation of neural network architectures, Springer-Verlag, London, 2002.
[64] Rosenblatt, F., The Perceptron: a probabilistic model for information storage and organisation in the brain, *Psychol. Rev.*, **1958**, 65(6): 386-408.

[65] Hopfield, J. J., Tank, D. W., Computing with neural circuits: a model, *Science*, **1986**, 233(4764): 625-633.
[66] Lewes, G. H., Problems of life and mind, Trübner & Co, London, 1875.
[67] Andersen, H., The history of reductionism versus holistic approaches to scientific research, *Endeavour*, **2001**, 25(4): 153-156.
[68] de Rosnay, J., Le Macroscope, Le Seuil, 1975.
[69] Lewin, R., Proposal to sequence the Human Genome stirs debate, *Science*, **1986**, 232(4758): 1598-1600.
[70] Lewin, R., Shifting sentiments over sequencing the Human Genome, *Science*, **1986**, 233(4764): 620-621.
[71] Roberts, L., Watson may head Genome Office, *Science*, **1988**, 240(4854): 878-879.
[72] Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J.-F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J. D., *et al.*, Whole-genome random sequencing and assembly of Haemophilus influenzae Rd, *Science*, **1995**, 269(5223): 496-521.
[73] Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., *et al.*, Initial sequencing and analysis of the human genome, *Nature*, **2001**, 409(6822): 860-921.
[74] Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., *et al.*, The sequence of the human genome, *Science*, **2001**, 291(5507): 1304-1351.
[75] Hood, L., Whitman College commencement address: "The Book of Life", Sunday, May 19, 2002, available at http://www.systemsbiology.org/.
[76] Guttmacher, A. E., Collins, F. S., Genomic medicine - A primer, *New England J. Med.*, **2002**, 347(19): 1512-1520.
[77] Weed, H. G., Medow, M. A., Guttmacher, A. E., Collins, F. S., Genomic medicine [4] (multiple letters), *New England J. Med.*, **2003**, 348(8): 759-760.
[78] Ksiazek, T. G., Comer, J. A., Rollin, P. E., Erdman, D., Peret, T., Emery, S., Tong, S., Roca, P., Bellini, W. J., Anderson, L. J., Goldsmith, C. S., Zaki, S. R., Humphrey, C. D., Shieh, W.-J., Guarner, J., *et al.*, A novel coronavirus associated with severe acute respiratory syndrome, *New England J. Med.*, **2003**, 348(20): 1953-1966.
[79] Kuiken, T., Fouchier, R. A. M., Schutten, M., Rimmelzwaan, G. F., Van Amerongen, G., Van Doornum, G., Osterhaus, A. D. M. E., Van Riel, D., Laman, J. D., De Jong, T., Lim, W., Ling, A. E., Chan, P. K. S., Tam, J. S., Zambon, M. C., *et al.*, Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome, *Lancet*, **2003**, 362(9380): 263-270.
[80] Rota, P. A., Oberste, M. S., Monroe, S. S., Nix, W. A., Campagnoli, R., Icenogle, J. P., PeÒaranda, S., Bankamp, B., Maher, K., Chen, M.-H., Tong, S., Tamin, A., Lowe, L., Frace, M., Chen, Q., *et al.*, Characterization of a novel coronavirus associated with severe acute respiratory syndrome, *Science*, **2003**, 300(5624): 1394-1399.
[81] Van De Vijver, M. J., Van 'T Veer, L. J., Voskuil, D. W., Peterse, J. L., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., Hart, A. A. M., Bartelink, H., Rodenhuis, S., Van Der Velde, T., Rutgers, E. T., Bernards, R., He, Y. D., *et al.*, A gene-expression signature as a predictor of survival in breast cancer, *New England J. Med.*, **2002**, 347(25): 1999-2009.

[82] Priori, S. G., Napolitano, C., Bloise, R., Ronchetti, E., Grillo, M., Vicentini, A., Nastoli, J., Bottelli, G., Folli, R., Cappelletti, D., Schwartz, P. J., Spazzolini, C., Risk stratification in the long-QT syndrome, *New England J. Med.*, **2003**, 348(19): 1866-1874.

[83] Yamada, Y., Tanaka, M., Izawa, H., Ichihara, S., Yokota, M., Takatsu, F., Ishihara, H., Hirayama, H., Sone, T., Prediction of the risk of myocardial infarction from polymorphisms in candidate genes, *New England J. Med.*, **2002**, 347(24): 1916-1923.

[84] St George-Hyslop, P. H., Petit, A., Molecular biology and genetics of Alzheimer's disease, *Comptes Rendus - Biologies*, **2005**, 328(2): 119-130.

[85] http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=gene.

[86] Simoes do Couto, F., De Mendonca, A., Garcia, C., Rocha, L., Lechner, M. C., Age of onset in patients with Alzheimer's disease with different apoE genotypes [1], *J. Neurol. Neurosur. Psy.*, **1998**, 64(6): 817.

[87] Ashford, J. W., APOE genotype effects on Alzheimer's disease onset and epidemiology, *J. Mol. Neurosci.*, **2004**, 23(3): 157-165.

[88] Houlden, H., Crook, R., Backhovens, H., Prihar, G., Baker, M., Hutton, M., Rossor, M., Martin, J. J., Van Broeckhoven, C., Hardy, J., ApoE genotype is a risk factor in nonpresenilin early-onset Alzheimer's disease families, *Am. J. Med. Gen.-Neuropsy. Gen.*, **1998**, 81(1): 117-121.

[89] Gygi, S. P., Aebersold, R., Absolute quantitation of 2D protein spots, *Meth. Mol. Biol.*, **1999**, 112(417-421.

[90] Futcher, B., Latter, G. I., Monardo, P., McLaughlin, C. S., Garrels, J. I., A sampling of the yeast proteome, *Mol. Cell. Biol.*, **1999**, 19(11): 7357-7368.

[91] Canguilhem, G., Le normal et le pathologique, Presses Universitaires de France, Paris, 1966.

[92] Collins, F. S., Green, E. D., Guttmacher, A. E., Guyer, M. S., A vision for the future of genomics research, *Nature*, **2003**, 422(6934): 835-847.

[93] Petricoin, E. F., Ardekani, A. M., Hitt, B. A., Levine, P. J., Fusaro, V. A., Steinberg, S. M., Mills, G. B., Simone, C., Fishman, D. A., Kohn, E. C., Liotta, L. A., Use of proteomic patterns in serum to identify ovarian cancer, *Lancet*, **2002**, 359(9306): 572-577.

[94] Petricoin, E. F., Mills, G. B., Kohn, E. C., Liotta, L. A., Proteomic patterns in serum and identification of ovarian cancer - Reply, *Lancet*, **2002**, 360(9327): 170-171.

[95] Pearl, D. C., Proteomic patterns in serum and identification of ovarian cancer, *Lancet*, **2002**, 360(9327): 169-170.

[96] Rockhill, B., Proteomic patterns in serum and identification of ovarian cancer, *Lancet*, **2002**, 360(9327): 169-169.

[97] Elwood, M., Proteomic patterns in serum and identification of ovarian cancer, *Lancet*, **2002**, 360(9327): 170-170.

[98] Diamandis, E. P., Proteomic patterns in serum and identification of ovarian cancer, *Lancet*, **2002**, 360(9327): 170-170.

[99] Petricoin, E. F., Ornstein, D. K., Paweletz, C. P., Ardekani, A., Hackett, P. S., Hitt, B. A., Velassco, A., Trucco, C., Wiegand, L., Wood, K., Simone, C. B., Levine, P. J., Linehan, W. M., Emmert-Buck, M. R., Steinberg, S. M., *et al.*, Serum proteomic patterns for detection of prostate cancer, *J. Natl. Cancer Inst.*, **2002**, 94(20): 1576-1578.

[100] Sorace, J. M., Zhan, M., A data review and re-assessment of ovarian cancer serum proteomic profiling, *BMC Bioinformatics*, **2003**, 4(1): 24.

[101] Baggerly, K. A., Morris, J. S., Coombes, K. R., Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments, *Bioinformatics*, **2004**, 20(5): 777-U710.
[102] Baggerly, K. A., Edmonson, S. R., Morris, J. S., Coombes, K. R., High-resolution serum proteomic patterns for ovarian cancer detection, *Endocr.-Relat. Cancer*, **2004**, 11(4): 583-584.
[103] Baggerly, K. A., Morris, J. S., Edmonson, S. R., Coombes, K. R., Signal in noise: Evaluating reported reproducibility of serum proteomic tests for ovarian cancer, *J. Natl. Cancer Inst.*, **2005**, 97(4): 307-309.
[104] De Veaux, R. D., Hand, D. J., How to lie with bad data, *Stat. Sci.*, **2005**, 20(3): 231-238.
[105] Check, E., Proteomics and cancer - Running before we can walk?, *Nature*, **2004**, 429(6991): 496-497.
[106] Koomen, J. M., Shill, L. N., Coombes, K. R., Li, D. H., Xiao, L. C., Fidler, I. J., Abbruzzese, J. L., Kobayashi, R., Plasma protein profiling for diagnosis of pancreatic cancer reveals the presence of host response proteins, *Clin. Cancer Res.*, **2005**, 11(3): 1110-1118.
[107] Diamandis, E. P., van der Merwe, D. E., Plasma protein profiling by mass spectrometry for cancer diagnosis: Opportunities and limitations, *Clin. Cancer Res.*, **2005**, 11(3): 963-965.
[108] DiMasi, J. A., Hansen, R. W., Grabowski, H. G., The price of innovation: new estimates of drug development costs, *J. Health. Econ.*, **2003**, 22(151-185.
[109] Baker, M., In biomarker we trust?, *Nat. Biotechnol.*, **2005**, 23(3): 297-304.
[110] Fleming, T. R., DeMets, D. L., Surrogate end points in clinical trials, *Ann. Intern. Med.*, **1996**, 125(7): 605-613.
[111] Butcher, E. C., Berg, E. L., Kunkel, E. J., Systems biology in drug discovery, *Nat. Biotechnol.*, **2004**, 22(10): 1253-1259.

Chapter 2. State of the art in the hyphenation of microfluidic devices with electrospray mass spectrometry

1. Introduction

Analytical biochemistry is undergoing a general trend toward miniaturisation, due to expected better analytical performances (e. g. in terms of sensitivity, time-to-result, and throughput), lower solvent and sample consumption, and disposability [1, 2]. This continuous effort has begun way before the advent of microfluidics, with downsizing of HPLC columns from 4.16 mm state-of-the-art packed columns down to 50 µm nanocolumns [3, 4]. With the emergence of microfluidics, analytical scientists could benefit from decades of developments in the microelectronics and microtechnology industry [5, 6] to build an entirely new generation of analytical devices [2, 7-15] that comprises DNA microarrays [16-18], high speed electrokinetic separations [19-23], biosensors for affinity sensing [24], immunoassays [25-28]… Because of their relative ease of implementation, most detection systems were based on optical techniques. However, the pioneering works of Ramsey's [29] and Karger's [30, 31] groups have dragged a lot of attention to the hyphenation of microfluidic devices with electrospray mass spectrometry [32-34]. The primary goal was to bring the analytical power of mass spectrometry to previously developed on-chip separations that reached very high efficiencies (from ten to hundred

thousands plates with low to sub-micron plate heights, with plate generation rates up to hundreds per second [11]), with additional advantage of disposability, potential low cost and reduced sample and solvent consumption. At the same time, the flow rate range typically encountered in microfluidics (from nano- to microlitres per minute) made them perfect candidates for hyphenation with electrospray ionisation mass spectrometry. A relatively straightforward way to hyphen microfluidic devices with mass spectrometry is to couple the microchip with a classical electrospray source, or to directly paste a micro- or nanospray device at its outlet. Though very successful, this approach suffers from a number of limitations, such as increased dead volumes, poor reproducibility and limited possibility of industrial production. A more difficult approach is to generate electrospray directly from the microchip, without any complementary transfer lines or pulled capillary nanosprays. Despite very successful proof-of-principle, designing a direct microfluidic interface for electrospray mass spectrometry turned out to be a difficult technical task; thanks to decades of developments in the microelectronics and microtechnology industry, almost any combination of size (down to a few microns or less), shape, material (originally silicon or glass but more recently also polymers) is available. It is thus of utmost importance to properly understand the main parameters that affect the formation of a stable electrospray from a microchannel, nozzle or capillary outlet. However, though it is known for long that capillary outlet size, flow rate and solvent composition can have dramatic effects on mass spectrometric signals, no clear scaling laws are established for electrospray ionisation, and most developments are still based on empiric rules and trial-and-error approaches. So after briefly reviewing the coupling of microfluidic devices with ESI-MS through classical electrospray sources or pasted nanosprays, some elements of electrospray theory pertinent to electrospray downscaling are highlighted, and we review herein all the technological and

analytical approaches for the coupling of microfluidic devices with electrospray mass spectrometry.

2. Hyphenation of microfluidics with ESI-MS through classical electrospray interfaces

The primary goal of microfluidics hyphenation with mass spectrometry being to take benefit of previously developed microchip separations with additional mass spectrometric detection, the most straightforward approach was to couple the microfluidic device with an existing electrospray ion source. Figeys *et al* introduced a multi-channel glass microchip for sequential sample delivery to ion-trap MS [35]: sample ports of the microchip could be automatically loaded with 2D-GE separated and digested proteins. Samples were then sequentially electroosmotically driven into a common microchannel connected to a classical microspray interface. This device allowed identification of proteins through MS/MS with limited cross contamination. Similarly, the same team developed an electrokinetic gradient generator [36]: the microchip had two reservoirs (one containing acidified acetonitrile, one containing acidified water) connected to a transfer capillary. A small solid-phase extraction (SPE) cartridge was first loaded with tryptic digests, and then connected in one extremity to the transfer capillary from the microchip and in the other extremity to a microsprayer. By varying the voltages applied to the two microchip reservoirs, authors were able to generate gradients, mainly driven by the transfer capillary acting as an electroosmotic pump. Upon increasing acetonitrile concentration, peptides eluted from the SPE cartridge, and were analysed by MS and MS/MS for protein identification. More recently, the same authors presented a more robust device including a filtration system in sample ports, and capillary coatings to kill the electroosmotic flow in transfer lines [37]. In a similar way, Meng *et al* tested the robustness and chemical compatibility of poly(methyl methacrylate) (PMMA)

microchip for coupling with Fourier Transform - Ion Cyclotron Resonance - Mass Spectrometer (FT-ICR-MS) [38]. In a more refined approach, Li *et al* directly connected a glass capillary electrophoresis microchip with a MicroIonSpray interface for the analysis of tryptic digests [39]; plate heights of 25000-600000 were obtained, which is less than state-of-the-art capillary electrophoresis (CE) with 90 cm glass capillaries, and it is highly probable than part of the separation took place in the fused silica capillary used to connect the chip to the MicroIonSource. Vrouwe *et al* also coupled a CE microchip with ESI-MS through a classical nanospray source for the analysis of small β-agonists [40]: though a simple three-components mixture was reproducibly separated with good sensitivity (nanomolar range on a single quadrupole instrument), the separation efficiency remained questionable, probably because the same voltage was used for the CE separation and for electrospray generation. Sample clean-up steps were also coupled to ESI-MS following this approach, such as on-chip microdialysis [41, 42] or free-flow electrophoresis to couple cIEF with mass spectrometric detection [43]. The most elaborated device coupled to ESI-MS through a classical electrospray interface was presented by Gao *et al* [44]: a PVDF membrane with immobilised trypsin was sandwiched in a PDMS microchip, and coupled through a transfer line to a microdialysis junction for direct injection into ESI-MS. Alternatively, transient cITP/CZE was performed downstream the microdialysis junction for separation of tryptic peptides. Though the separation efficiency may remain questionable, the approach provided a complete platform for protein digestion, peptide separation and analysis by ESI-MS.

In a more developed approach, Wachs *et al* introduced an electrospray interface dedicated to the hyphenation of microfluidic devices with electrospray mass spectrometry [45]. Shown in Figure 14 is a schematic of the liquid junction device, allowing the addition

of a makeup flow, a sheath gas, and a final electrospray from a 31 gauge (226 μm) stainless steel tubing.

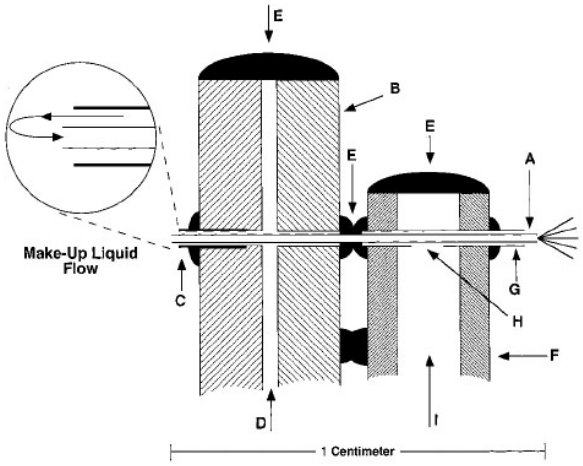


Figure 14. Micro ion spray source designed for interfacing with chip-based separations, incorporating a self-contained liquid junction. (A) 31 gauge stainless steel tubing, (B) 0.020 in. i.d. ×0.0625 o.d. stainless steel tubing, (C) and (G) 23 gauge stainless steel tubing, (D) makeup liquid entrance, (E) seal, (F) 0.0625 in Teflon tubing, (H) hole and (I) gas inlet. Reprinted with permission from [45].

This device was further used to hyphen a polymer capillary electrophoresis microchip with a triple quadrupole mass spectrometer for the detection of carnitines [46-48]; the electrospray voltage was directly applied to the microsprayer (typically 3 kV), and a higher voltage was applied at the upstream reservoir of the microchip (typically 5 kV); the separation was performed in pure aqueous solution. In a first demonstration, 0.3 pmol of carnitines were separated within 30 s, and detected both in full san mode and in single reaction monitoring. Optimisation of the separation conditions (separation field strength) resulted in quicker separations (down to ten seconds) with plate counts ranging from 4800 to 18000. The same micro-liquid junction was also used more recently to investigate P450 drug metabolism on a chip [49]: one flow containing cytochrome P450 and NADPH-cytochrome P450 reductase and one flow containing small drugs were mixed on-chip in a stopped flow manner, before the mixture was further desalted and analysed by triple

quadrupole thanks to the micro-liquid junction. Variation of the on-chip incubation time coupled to mass spectrometric detection allowed a straightforward determination of degradation products of the drugs studied. Complementarily, the same group used this micro-liquid junction to characterise acrylate-based monolithic solid-phase adsorbents synthesised on-chip [50]. Basically, this approach provides an elegant way to couple microfluidic devices to ESI-MS without the need for transfer capillary that usually alters separation efficiencies, and requires chemical treatments for electroosmotic flow suppression; moreover, sheath gas and make up flow can be tuned for optimal ionisability and desolvation of analytes of interest.

Another way to hyphen microchips with electrospray mass spectrometry is to glue micro- or nanospray needles at the outlet of a microchannel for direct electrospray generation. This approach was followed by numerous authors for simple sample delivery [51-54]; even without on-chip separation or sample clean-up, these devices already provide the benefit of loading the nanospray tip by electroosmotic flow instead of manual pipetting or centrifugation. Nanosprays were also pasted to microchips for on-chip flow injection analysis through a double T injector for the analysis of subattomole amounts of peptides and proteins [55, 56]. It is worth highlighting that as soon as voltages are used to move samples electrokinetically, proper decoupling of electrospray and "driving" voltages must be accomplished: Lazar *et al* proposed a glass membrane to apply electrospray voltage near the outlet of a glass microchip, without any further dilution of samples coming from the main microchannel, as is the case with a direct T junction. With this device authors were able to detect less than 4 attomoles of Gramicidin S on a TOF instrument, with a pasted microspray [56]. Direct pasting of nanospray tips at the outlet of microchannels was also used by Wu *et al* to couple a PDMS microreactor with immobilised trypsin with MS detection for protein identification [57].

But pasting micro- or nanosprays at the outlets of microfluidic channels has been mainly used to hyphen previously developed on-chip separations with ESI-MS. For example, Sung *et al* used a PDMS CE microchip coupled to ESI-Q-TOF to analyse model peptides, and tryptic digests [58]; though separation efficiencies were not thoroughly investigated, online MS/MS was demonstrated for protein identification. Similarly, Lazar *et al* developed a glass microchip for capillary electrochromatography coupled to ESI-MS [59]: a monolith phase based on methacrylate derivatives was photopolymerised within the separation microchannel and further chemically modified with N-ethylbutyl amine to control electroosmotic flow and provide hydrophobic interactions with peptides. Porous glass membranes in the reservoirs also allowed voltage delivery without parasite fluid movements. This device allowed analysis of tryptic digests on a TOF instrument in full scan mode.

Harrison's laboratory probably presented the most integrated and successful proofs-of-principle of the hyphenation of on-chip sample preparation and separation to ESI-MS through glued nanosprays; Figure 15 shows the generic design of the glass microchip used for CE separation of peptides and on-line sample delivery to ESI-MS:

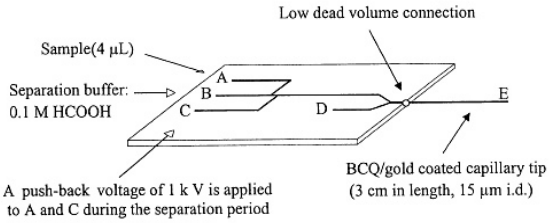


Figure 15. Design of the glass microchip used for CE separation and nanoelectrospray. Reprinted from [60] with permission.

Sample is loaded in microchannel A, then driven electrokinetically to microchannel C, the voltages in B and D being left floating. At a certain point, voltages are switched, so that sample migrates into the separation channel (pushback voltages being applied to A and

C to avoid sample leakage during separation). In parallel, a standard compound was infused continuously through microchannel D for mass calibration of the Qq-TOF instrument used. Model protein digests were analysed, with plate numbers ranging from 4100 to 7400 with plate heights of 9-17 μm and plate generation rates in the range 50-100 plates per second; moreover, peak capacity could be estimated to be in the range 20-25 from the electropherograms shown [60]. When combined with on-line sample stacking, this methodology reached sub-nanomolar limit of detection for peptides. This device was further applied to the identification of membrane proteins from *H. Influenzae* separated by 1D-GE [61]. In a more integrated approach, microchannel C shown in Figure 15 was made bigger to accommodate a microbead bed, whether for trypsin immobilisation and online protein digestion [62], or for solid-phase extraction on C18 material: in this case, the SPE bed could be sequentially loaded by an autosampler with tryptic digest of proteins from *N. meningitidis* separated by 1D-GE; peptides were then separated by on-chip CE and analysed by MS: 67 gel bands were analysed at a pace of 2 min/samples, resulting in 55 proteins identified and 5 additional gene variants identified [63]. In its most accomplished demonstration, the same device was used for identification of proteins from human prostatic cancer cells (LNCap) separated by 2D-GE: authors were able to analyse typically 200 fmol of in-gel protein digests in an automated way. Alternatively, they successfully replaced the C18 material by Immobilised Metal Ion Affinity Chromatography (IMAC) microbeads for isolation and enrichment of phosphorylated tryptic peptides prior to on-chip CE and MS analysis [64].

The main drawback of this approach is that nanosprays must be manually inserted into the exit microchannel and somehow pasted or sintered, which compromises the robustness, reproducibility and possibility of mass production of the approach. Tashibana *et al* presented an elegant alternative to hyphen a nanospray tip to a quartz CE microchip

[65]: the extremity of the microchip was machined so that it can accommodate a PEEK screw; a 370 µm O.D. nanospray tip could then be directly connected or replaced without any glue or sintering, while incorporating a liquid junction on the microchip to apply the electrospray voltage. In a further study with this device, separation conditions (separation channel length and coating) were optimised, so that a mixture of 19 amino acids, or a cytochrome C tryptic digest, could be separated and analysed by MS [66].

3. Microfluidic devices with direct electrospray generation

As discussed above, the hyphenation of microfluidic devices with ESI-MS through classical electrospray sources requires the use of capillary connections and transfer lines that complicate the control of electroosmotic flow, add dead volumes, usually requires sample dilution with sheath flows for efficient ionisation and desolvation, and may lower separation efficiencies through diffusion processes. The use of pasted micro- or nanosprays just at the outlet of the microchip addresses some of these drawbacks by diminishing dead volumes and usually increasing the sensitivity of the mass spectrometric detection compared to classical electrospray sources. However, reproducibility of nanospray ESI-MS analyses remains questionable, and the gluing process cannot be transferred to industrial scale. Direct generation of electrospray from a microchip without any additional device has thus been investigated by numerous teams using different materials for microchip production, different designs and analytical strategies.

3.1. Elements of electrospray theory

Electrospray establishment requires a well-defined solution/air interface at the outlet of a capillary or microchannel. Without any applied field, a meniscus establishes, the geometry of which is governed by the outlet geometry and the liquid surface tension. When

the electrospray voltage is applied to the solution, charges accumulate at the solution/air interface and the interface shape is then governed by the equilibrium between surface tension and electrostatic pressure. If the applied voltage is high enough, the interface takes the shape of a cone, the so-called Taylor cone. At a certain potential U_T, the apex of the Taylor cone is destabilised and charged droplets are emitted. The formation of gas-phase ions from these electrospray droplets was the subject of active debates. Two theories emerged, namely the *charge residue model*, in which electrospray droplets shrink and explode under solvent evaporation till only one ion sits at the centre of each droplet and is finally desolved, and the *ion evaporation model*, in which ions can be directly expelled from electrospray droplets through charge repulsion mechanisms [67]. It has been know since the very beginning of electrospray ionisation that for a given capillary size, the electrospray current (and thus the number of analyte ions detected by the mass spectrometer) is highly dependent on the flow rate, as shown by Pfeifer and Hendricks for spray in vacuum [68]:

$$i_{ES} = \left(\frac{4\pi}{\varepsilon}\right)^3 (9\gamma)^2 \varepsilon_0^5 (\kappa E)^{3/7} \nu^{4/7} \qquad (1)$$

where ε and ε_0 are the permittivity of the solvent and of vacuum respectively, γ the surface tension of the solvent, κ its conductivity, E the electric field and ν the solution flow rate. Similarly, de la Mora found that the electrospray current was proportional to the square root of the sample flow rate [69]. But in the early nineties, Emmett and Mann introduced in parallel micro- [70] and nano- [71] electrospray ionisation respectively; both approaches showed that the use of smaller capillaries (associated with lower solution flow rates) resulted in much higher ionisation efficiencies. The reason for that is that more and smaller electrospray droplets are formed when a smaller capillary is used, and that gas phase ion formation is much more favourable when starting from smaller droplets. The size of

electrospray droplets is thus of utmost importance in the final sensitivity of an electrospray interface. Unfortunately, their size (a few tens of nanometres to a few microns) and dynamics make them poorly amenable to observation. To our knowledge, no in-depth study has been conducted on the dependency of electrospray droplet size on capillary outlet dimension. Wilm and Mann introduced an estimate for the radius of the droplet emitting zone from relatively simple geometrical considerations [71]:

$$r_e = \left(\frac{\rho}{4\pi^2 \gamma \tan\left(\frac{\pi}{2} - \upsilon\right)\left[\left(\frac{U_a}{U_T}\right)^2 - 1\right]} \right)^{1/3} \nu^{2/3} \qquad (2)$$

where ρ is the density, υ is the Taylor cone angle, U_a is the applied electrospray voltage, U_T is the threshold voltage at which electrospray appears, and all others as in Equation (1). Figure 16 shows the value of the droplet-emitting zone radius versus the applied flow rate for water (ρ = 1000 kg/m^3, γ = 72.8 10^{-3} N/m) in typical electrospray conditions (U_T = 1000 V, U_a = 1500 V).

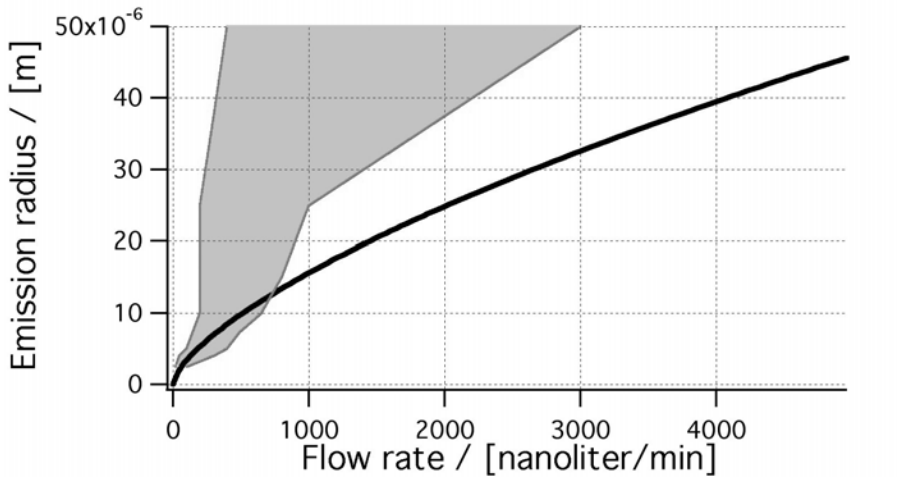


Figure 16. Size of the droplet-emitting zone versus the applied flow rate (solid line).

This quantity is obviously a rough overestimate of droplet radius, but qualitatively gives an indication of the outlet diameter necessary to generate a stable electrospray at a

given flow-rate, as shown by comparison with experimental results [71]. Interestingly, this estimate is in qualitatively good agreement with empirically recommended flow rates (grey zone in Figure 16, from [72]), especially for low flow rates.

3.2. Different design approaches

Influence of material

The role of the capillary or microchannel outlet is to confine the liquid to be sprayed to a definite size and geometry, so that a stable Taylor cone can establish under the application of the electrospray potential. This implies that the spray solution should not wet the outer surface of the capillary or microchannel. The hydrophobicity of the material used to build the microfluidic electrospray emitter is thus of utmost importance; for example, Xue *et al* found it difficult to generate electrospray directly from the edge of a glass microchip such as the one shown in Figure 17.

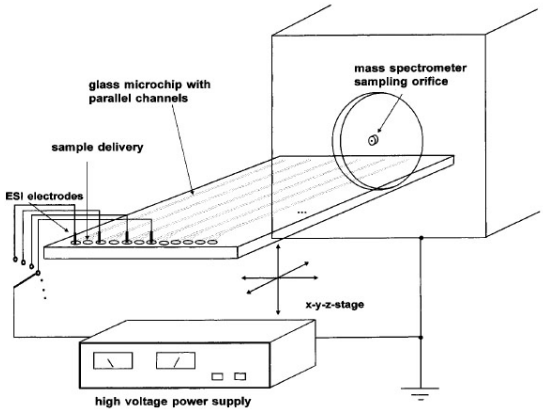


Figure 17. Scheme of the glass microchip and set-up used by Xue et al *in [31].*

In fact the spray solution had tendency to moisten the edge surface of the chip, which prevented any stable spray. The authors had to coat the edge of the chip with a hydrophobic reagent such as n-octyltriacetoxysilane to avoid any wetting of the edge surface of the chip and obtain a stable electrospray. On the contrary, Ramsey *et al* did not

find any improvement in spray operation when silanising the microchannel outlet of a very similar glass microchip [29]; this difference may be due to differences in glass material, and surface state after microtechnological processing. However, at least qualitatively speaking, it is necessary to choose a material that is hydrophobic enough to avoid wetting of the edge surface (or nozzle outer surface). Whereas the values presented in Figure 18 may depend a lot on the microtechnological processing (exposure to plasmas, oxidising agents…), it clearly shows that polymer materials are usually much more hydrophobic than glass or silicon dioxide. They thus appear as good candidates for electrospray emitter design.

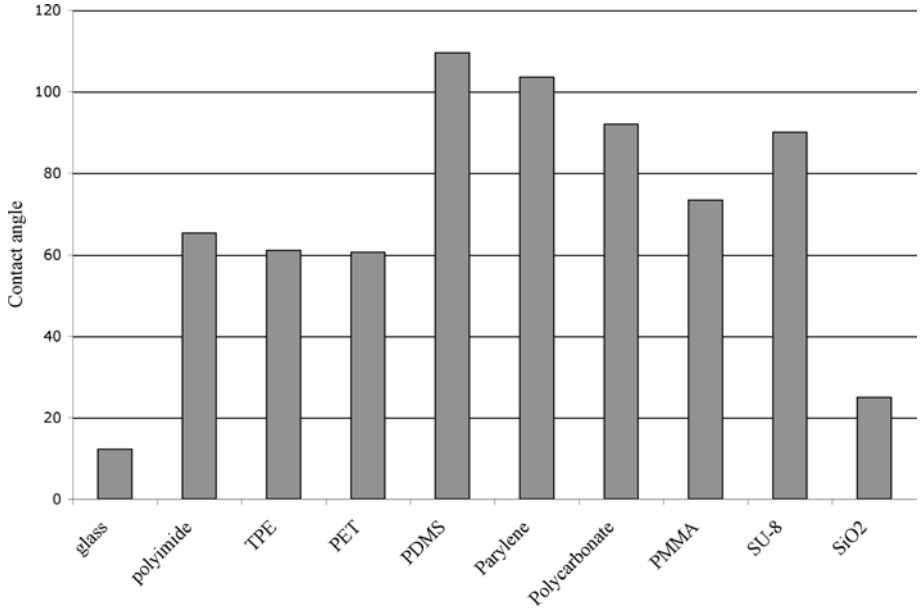


Figure 18. Static contact angle measurements of various materials.

Another way to properly design the outlet wall hydrophobicity is to modify the surface properties by a chemical treatment: as mentioned above, Xue *et al* used a n-octyltriacetoxysilane coating, but found it relatively unstable under operation [31]. Tang *et al* used a CF_4 radio frequency plasma treatment to increase the hydrophobicity of polycarbonate electrospray emitters [73], but rf plasma treatments are usually of short longevity through exposure to air or aqueous solutions. Wang *et al* introduced recently an elegant way to modify the electrospray outlet hydrophobicity on polymer electrospray

emitters [74]: a poly-tetrafluoroethylene (PTFE) hydrophobic membrane is pasted over the outlet polycarbonate surface of an electrospray emitter. Due to this surface modification, very stable electrosprays were established at flow rates as low as 10 nL/min.

Importance of electrode design

In order to generate electrospray, high voltage has to be applied to the solution to be analysed. Charge separation then occurs within the electrospray emitter, leading to electrospray droplet generation. The electrospray set-up can thus be described in terms of electrical circuit [75], as depicted in Figure 19.

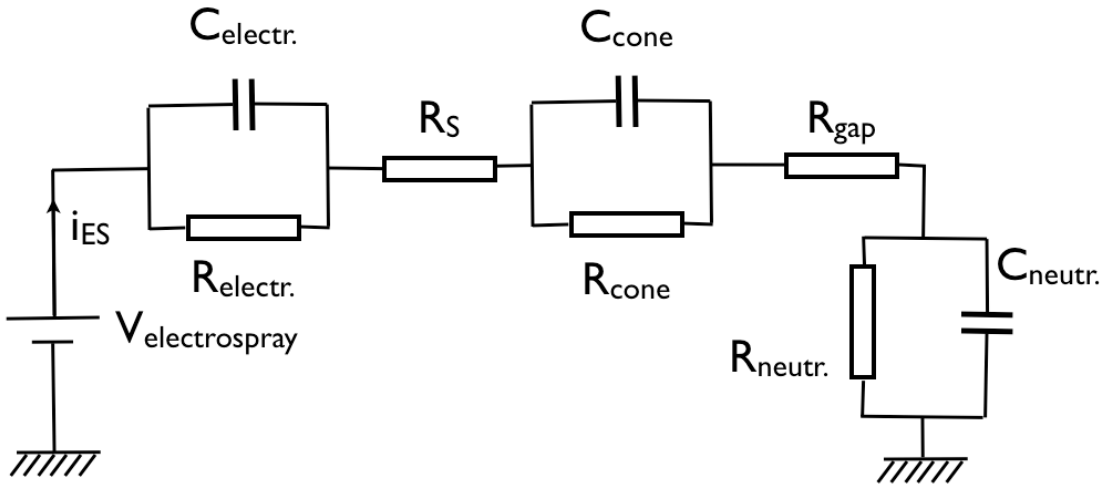


Figure 19. Electrical circuit equivalent to the electrospray process. $V_{electrospray}$ *is the potential applied on the electrospray electrode,* $C_{electro.}$ *and* $R_{electr.}$ are the capacity and charge transfer resistance of the electrospray electrode, R_S *the solution resistance,* C_{cone} *and* R_{cone} *are the capacity and charge transfer resistance of Taylor cone where charged analytes accumulate,* R_{gap} *is the resistance of the air gap between the apex of the Taylor cone and the mass spectrometer inlet,* $C_{neutr.}$ and $R_{neutr.}$ *are the capacity and charge transfer resistance of the detector where ions are neutralised.*

The electrospray emitter can be viewed as a particular type of controlled-current electrochemical cell [76] in which the electrode potential auto-adjusts to sustain the

electrospray current dictated by the emission rate of charged droplets at the Taylor cone apex: in positive ion mode, the electrode potential will rapidly increase up to the oxidation of the first oxidable specie; at this point, the current will be limited by the mass transport of this specie to the electrode and remain stable. When this specie is fully consumed, the electrode potential will increase again up to the oxidation potential of the second oxidable specie and remain constant. Only in the case of solvent oxidation is the current not limited by mass transport to the electrode.

In most cases, one wants to avoid any electrochemical reaction of analytes of interest on the electrospray electrode. It is thus of particular importance to properly design the electrode material, size and position respective to the Taylor cone. In this instance, it is obviously desirable to choose an inert material for the electrode, such as platinum, gold or graphite. Moreover, for a given electrospray potential applied and a given microchannel outlet/mass spectrometer distance (i.e. for a given current), the current density through the electrode, and thus the kinetics of the electrochemical reactions at the electrode surface will be defined by the electrode effective area; more precisely, reaction rates are inversely proportional to the electrode effective area [77], which means that the use of larger electrodes results in globally lower electrochemical reactions. Lastly, when reactive species are potentially created at the electrospray electrode, their residence time within the electrospray emitter determines how probably they can react in the homogeneous phase with analytes of interest. Placing the electrospray electrode as close as possible to the Taylor cone can thus limit adverse electrochemical degradation of analytes.

3.3. The off-axis, nozzle-like approach

Microtechnological processes rely essentially on planar technologies, where substrates are perpendicularly exposed to light or etching agents through masks, or lasers are scanned over the substrate surface to produce microfeatures [5, 6]… Most simple

processes thus allow only to produce holes or microchannels through a substrate, perpendicular to its surface (off-axis design), or features patterning on a planar surface, parallel to its surface (on-axis design). Though historically the most recent approach, the off-axis strategy has given rise to the first commercially available product from Advion: the ESI-Chip™ [78]. It basically consists in a silicon substrate in which arrays of nozzles are etched by a relatively complex process (around fifteen elementary operations are listed in [78], not including the final surface treatments). The elementary nozzle, shown in Figure 20, is a 10 µm i.d., 20 µm o.d., 50 µm long silicon cylinder; arrays of 100, 200 and 400 nozzles are commercially available.

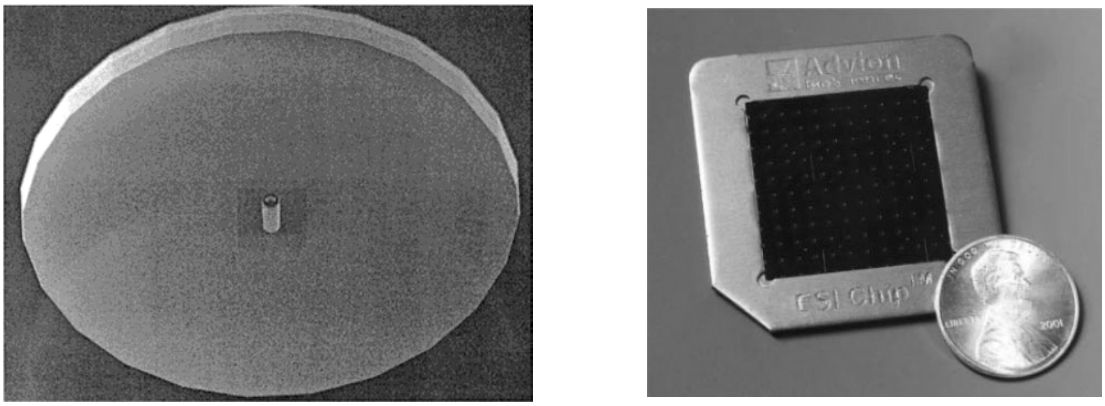


Figure 20. Left: scanning electron microscope image of an Advion nozzle (10 µm i.d., 20 µm o.d.); reprinted from [78] with permission. Right: the ESI-Chip comprising 100 nozzles (reprinted from [79] with permission)

The ESI-Chip™ is operated thanks to a dedicated instrument, the Nanomate™, which basically pipets the sample from a 96 well plate with a single-use conductive pipette tip; the pipette is then pushed against the back side of the ESI-Chip™ and the sample is pneumatically pushed through the nozzle, at typical flow rates between 50 and 300 nL/min. The electrospray voltage (typically 800-1500 V) is applied to the solution through the conductive pipette tip, while the front side of the ESI-Chip™ is grounded. In its initial demonstration [78], the device was shown to allow the analysis of pure aqueous samples with better stability (2.1% RSD vs 9.8% RSD over five minutes) and sensitivity than pulled

capillaries of comparable dimensions. Moreover, nozzle-to-nozzle reproducibility was evaluated to be 12.1% RSD within 10 nozzles. In a second development stage, nozzles of 8 µm i.d., 30 µm o.d. were produced.

The ESI-Chip™ in combination with its robotic system, the Nanomate™ was interfaced to a wide variety of mass spectrometers, including time-of-flight [78, 80-85], ion traps [86, 87], triple quadrupoles [79, 88-91] and Fourier Transform- Ion Cyclotron Resonance- Mass Spectrometer (FT-ICR-MS) [92], in both positive [78, 79, 81, 82, 86-91] and negative [83-85, 87, 92] ionisation polarity, and for simple MS as well as more complex experiments, such as in-source fragmentation [82], selected reaction monitoring (SRM) [79, 88-91], collision induced dissociation [83-87] and sustained off-resonance irradiation collision induced dissociation [92].

The advantages of the automated Nanomate™ system are clear: first it allows automated nanospray sample delivery to the mass spectrometer. For example, Van Pelt *et al* showed the quantitation of drugs from Caco-2 permeation assay [79]: the duty cycle used was 100 s (40 s for sample handling and 1 min of MS analysis), which corresponds to 2h40 for the analysis of 96 samples. Interestingly diminishing the MS analysis time to 5 s would make 72 min for the analysis of 96 samples. Kapron *et al* even claims the possibility of analysing 120 samples per hour with this system [89]. As such, the Nanomate™ system can be seen simply as a high-throughput sample delivery system to electrospray ionisation mass spectrometry. It is particularly interesting when coupled off-line to upstream separations for the analysis of complex samples. For example, Van Pelt *et al* used the system for the identification of proteins separated by 2D-GE, stained by Comassie blue or silver, excised from the gel, digested by trypsin and solubilised [86]: 140 proteins from a crude cell extract of *Saccharomyces cerevisiae* were identified with a duty cycle of 210 s per sample (8h10 of unattended analysis). This study highlighted the possibility to use the Nanomate™ system

just as MALDI ionisation would be used, with additional benefit of being interfaced to an ion trap allowing tandem MS at high speed. Similarly, Bindila *et al* used off-line capillary electrophoresis coupled to the Nanomate system for sample delivery to a Q-TOF instrument for the analysis of complex carbohydrates mixtures [84]. This kind of analytes is particularly challenging because they are poorly ionisable. In this study, analytes could be directly analysed in the CE buffer thanks to the quality of the ESI-Chip nozzles, and MS/MS experiments could be conducted for structural elucidation of carbohydrate compounds without being hampered by the CE pace. Complementarily, Staack *et al* used classical online LC/MS/MS while collecting part of the LC effluent for off-line interrogation through the ESI-Chip of specific parts of the chromatogram and structural analysis of minor components, in the context of drug metabolism studies [93].

Nozzle quality and electrospray conditions optimisation are known to be crucial for the analysis of non-covalent complexes [94]; the Advion analytical platform has been successfully applied to the qualitative screening of ligands to fatty acid binding proteins, with analysis throughput of 430 samples per working day and protein consumption around 100 pmol per ligand assay [95]. Keetch *et al* also demonstrated the usefulness of the ESI-Chip for the analysis of large protein multimers and qualitative investigation of inhibitor binding to these large multi-protein complexes [82]. More interestingly, quantitative information could be obtained about the binding of small ligands to proteins: De Vriendt *et al* showed that the binding constant between cellobiohydrolase I and cellulose as well as inhibitors could be measured by direct visualisation of non-covalent complexes in the mass spectrum [96], and Zhang *et al* similarly calculated the binding constants for different systems, such as RNAse A with monophosphate nucleotides [81].

But the nozzle quality may also allow to completely suppressing upstream sample separation. For example, Kapron *et al* were able to quantify small drugs from human

plasma with very limited sample preparation and no separation at all, whereas the state-of-the-art method consists in LC/MS/MS [89], as well as Dethy *et al* in a similar study [90]. This approach has even been validated recently in the context of Food and Drug Administration regulatory guidelines and Good Laboratory Practices [97]. The benefit of eliminating the LC separation for certain applications is clear: first the throughput depends only on sample delivery to the MS and not on separation efficiency and pace. The extension of this approach outside the drug quantitation field will probably require more mass spectrometry developments such as the ones presented by Kalkum *et al* for the detection of target peptides from crude biological extracts [98], but it could lower the demand on upstream separation efficiencies that is currently seen in proteomics for example.

Interestingly, the laboratories of Stemme and Roeraade followed an approach similar to that of Advion for the design of silicon-based arrays of electrospray nozzles [99]. They tested two different nozzle designs, based whether on pure silicon or on silicon dioxide. Figure 21 shows the two types of nozzles:

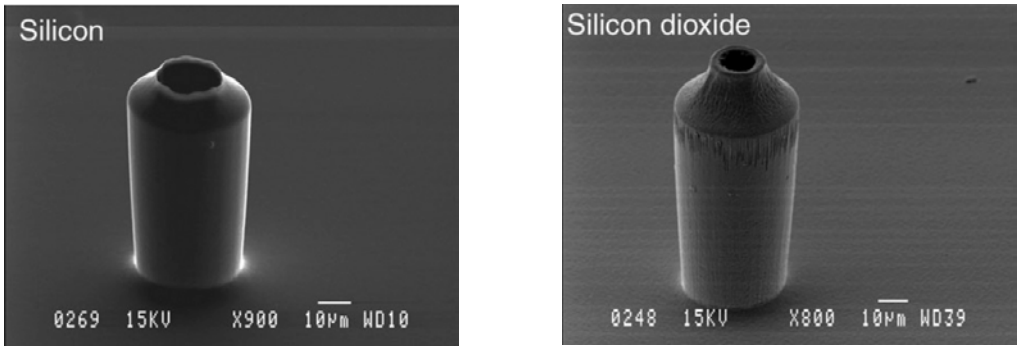


Figure 21. Silicon and silicon dioxide nozzles. The silicon nozzle consists in a straight cylinder hole with wall of variable thickness, whereas the silicon dioxide nozzle is made of a 1.5 µm layer of silicon dioxide of constant thickness. Reprinted with permission from [99].

The one made of silicon consists in a straight cylinder with wall of variable thickness, and the one in silicon dioxide is made of a structure of constant thickness (the internal shape is the same as the external one). The main difference with ESI-Chips from

Advion is that for both designs, the outlet has a very sharp surface, whereas in ESI-Chips, the nozzle outlet has a flat surface (see Figure 20). Silicon dioxide nozzles were further tested for peptides, proteins and non-covalent complexes analysis [100].

3.4. The planar, nozzle-less approach

As discussed above, most microtechnological processes allow microprocessing through substrates (off-axis design), which Advion used to produce the ESI-Chip, or microprocessing on the substrate surface (on-axis design), parallel to its plane, to produce microchannels. Historically, this kind of devices has been developed first for on-chip separations [19-22]. The most straightforward way to couple such devices to electrospray mass spectrometers was to cut the microchip perpendicular to the microchannel, so that the microchannel outlet is located on the edge of the chip. This solution was pioneered by Ramsey's [29, 101] and Karger's [30, 31] groups. As mentioned in paragraph "Influence of material", it rapidly appeared that the electrospray stability was somehow compromised with such an interface. But several teams simultaneously introduced a triangular-shaped microspray emitter [102-107], a design that was rapidly adopted by other groups. The rationale for a better stability of electrospray in such designs was never fully explained, but one can reasonably argue that having a microchannel outlet just at the extremity of a triangular microchip (instead of a flat edge) results in a better confinement of the solution / air interface by limiting the wetting of microchip outer walls by the solution. A lot of different designs have been proposed on PDMS [103-105], PET [107], polyimide [108], PMMA [106, 109], polyester [110]… Some other groups worked on alternate designs: Licklider *et al* compared "blunt tips" with oval apertures and "sharp tips" made of parylene, as shown in Figure 22;

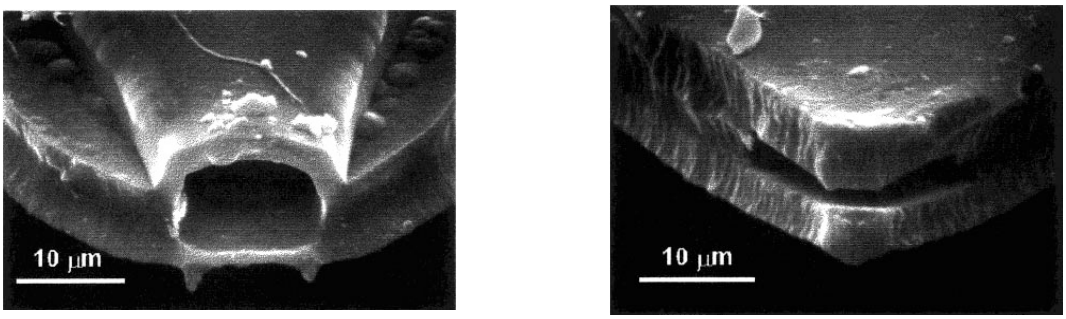


Figure 22. Parylene "blunt tip" (left) and "sharp tip". Reprinted with permission from [102].

Basically both devices showed the same analytical properties as a classical pulled capillary microsprayer. Alternatively, Kameoka *et al* sandwiched a parylene triangle between the two plastic layers forming the microfluidic chip, so that the liquid coming out from the microchannel wetted the parylene triangle, at the tip of which electrospray was generated [111]. This device was further used for methylphenidate (Ritalin) quantitation from human plasma [112].

Most designs were demonstrated as stand-alone devices for sample delivery to ESI-MS, whether under infusion at flow rates from 0.1-1 μL/min, or with reservoirs under electroosmotic flow. Interestingly, Lion *et al* showed that polyimide microspray emitters with half moon cross sections of 120×45 μm can be used from 250 nl/min up to 5 or 7 μl/min in a wide range of solvent compositions with better stabilities and comparable sensitivities than pulled capillary nanospray, whose applicable flow rate range is much more limited for a given aperture size (see Figure 16) [113]. This is somehow contradictory to the general consensus that the smaller the electrospray emitter, the more sensitive the analysis. The same device was applied to the analysis of proteomic samples [114], and glycoconjgate screening [115] when coupled with FT-ICR-MS or with Q-TOF-MS [116].

But other analytical functions can also be performed on-chip prior to MS analysis, such as solid-phase extraction (SPE) for sample desalting and clean-up: Lion *et al* placed a

small piece of PVDF membrane at the inlet of the microsprayer to capture drugs, peptides and proteins, that can be further washed and eluted by the spraying solution [117, 118]. More interestingly, Rohner *et al* introduced a new concept for online electrochemical tagging of cysteine residues [119]: a neutral probe from the hydroquinone family is introduced with the proteinaceous sample to analyse; hydroquinone being the first oxidable specie, it is converted into benzoquinone on the carbon-ink electrospray electrode, which is selectively reactive toward free cysteine residues via a Michael addition. In conditions where this mass-tagging is non-quantitative, the presence of each free cysteine in the analyte is revealed by an adduct peak in the mass spectrum, and thus a direct counting of the number of cysteines in a given peptide. After elucidation of the electrochemical mechanism [120], optimisation of the microchip design [121] and of the electrochemical probe reactivity [122], this process was applied to protein identification through peptide mass fingerprinting [123]: when the number of cysteine is added to the peptide masses during database interrogation, the identification score is greatly increased, thus providing more confidence in protein identification. Though this process might be equally feasible with classical electrospray emitter on the principle, microfabrication provides unique possibilities of adjusting the electrode size and position, and the microchannel dimensions so that the overall electrochemical conversion rate of the probe on the electrode, and contact time between the converted probe and analytes of interest can be controlled [124].

3.5. The planar, nozzle-like approach

Whereas the planar, nozzle-less approach proved to be successful, alternative planar devices were introduced to mimic nozzles at the outlet of microchannels. For example, Wen *et al* early introduced a microchip with a cone-shape electrospray emitter made of polycarbonate [125]: this device was used to perform isoelectric focusing of proteins in broad range carrier ampholytes (pH 3-10). Once focused, proteins were chemically

mobilised while a sheath liquid was provided close to the electrospray emitter through a side microchannel. The same approach was followed by Svedberg *et al* to design hand-polished or machined-drilled cone-shaped electrospray emitters in PMMA microchips [126]. More recently, the same group designed an opened PDMS electrospray tip with a U profile for simple sample delivery through capillary aspiration, and demonstrated its use with model proteins [127]. Similarly, Rolando's group designed micro-nib electrospray sources, first made of SU-8 photoresist [128-132] and later of poly-silicon [133, 134], as shown in Figure 23. Basically, the sample solution is deposited in the nib reservoir and flows toward the nib tip where the electrospray is generated [135].

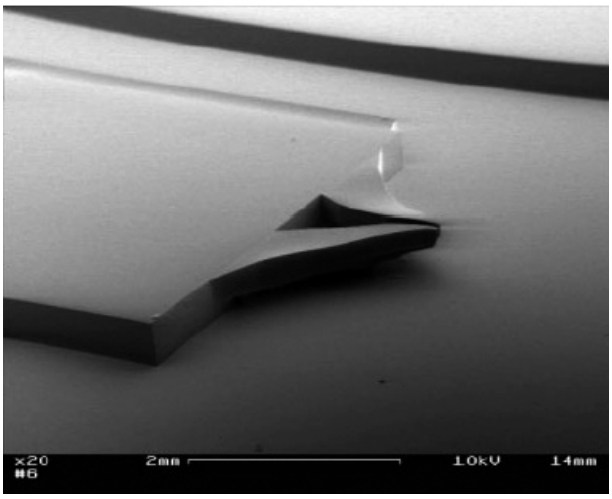


Figure 23. Nib-like electrospray emitter, made of SU-8 photoresist. Reprinted from [128] with permission.

The most comprehensive and successful planar, nozzle-like microfabricated electrospray emitter has been introduced recently by Agilent for on-chip HPLC/MS: three polyimide layers are first photoablated to produce microchannels, sample ports and restrictions for further microchannel packing. Electrical contacts are then deposited and the three layers are assembled to form the HPLC-chip. The end of the main microchannel is then laser-machined in a conical shape, with 35 to 100 µm outer diameter at its extremity (the microchannel outlet having typical dimensions of 5 µm), as shown in Figure 24.

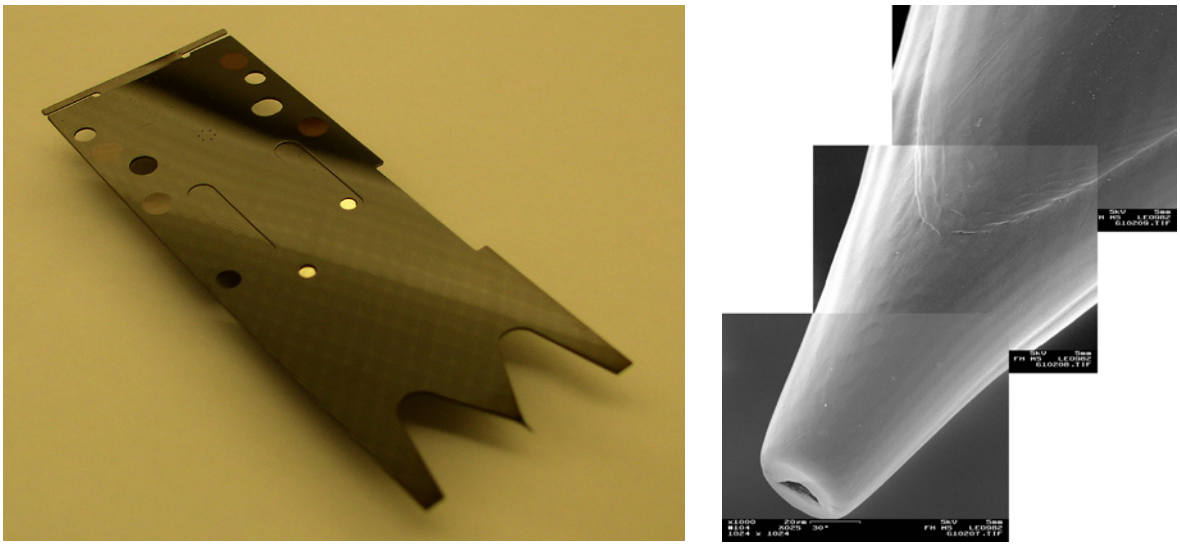


Figure 24. Left: Naked polyimide HPLC-MS microchip from Agilent, showing the inlet ports for the valving system, the electrical contacts and the electrospray tip. Right: magnification of the conical electrospray tip. Courtesy of Agilent Technologies, R&D and Marketing, Waldbronn, Germany.

The sample enrichment and separation microchannels are then packed with reverse phase microparticles, and the chip is mounted between the stator and the rotor of a two-position HPLC rotary switching valve [136]. In the initial demonstration, 20 fmol of BSA digest were analysed at 100 nL/min with satisfactory retention times and peak areas reproducibility. It is noteworthy that the microchip design was supported by detailed study of chromatographic dynamics in non-cylindrical microchannels [137]. In a second study, the same device was validated with more realistic samples: with an eight model protein mixture digest, peak capacity of 141 was observed over a 50 minutes gradient together with decent peak symmetries and excellent reproducibilities (average RSD values of 0.5% on retention time, 0.003 % on m/z and 9.1% on peak intensities over 10 replicates of the separation of 626 peptides). Authors also checked the potential of the device in the context of expression profiling by spiking 0 to 250 fmol of the same eight proteins digest in 100 ng digested rat plasma. Whereas at 1.2 fmol spiking, only four peptides among the expected 626 were detected, the number of identified peptides reached 604 for spike levels of 120

fmol. Moreover, the peptide intensities showed very good linearities in the tested range. Finally, rat plasma was analysed whether by 1D HPLC-MS (55 minutes gradient) or by 2D-HPLC-MS by inserting cation exchange material into the sample inlet port of the switching valve (eight salt fractions with 55 minutes reverse phase gradients). 47 nonreduntant proteins (111 in 2D-HPLC-MS) were readily identified with detection limits down to less than 80 ng/ml. Moreover, sequence coverages were two to three times higher in 2D-HPLC-MS [138]. More recently, a similar HPLC-MS microchip was used in the context of oligosaccharides analysis [139]: Ninonuevo *et al* replaced the reverse phase packing material by a graphite porous carbon phase, validated the device on model oligosaccharides and analysed O-linked oligosaccharides from mucins and oligosaccharides mixtures from milk. Authors obtained time retention variations of less than 0.5% on 40 minutes gradients, and peak area reproducibilites in the order of 5%. Moreover, the high mass accuracy of the TOF analyser used allowed direct assignment of most structures observed. Though this device is still at the validation stage, it seems that the unique combination of resources in microtechnological processing and mass spectrometry on an industrial scale resulted in a very promising analytical microdevices family; though the use of microchips did not result in this case in a dramatic increase of analytical performances (whether in sensitivity or in analysis time) compared to classical nanoLC columns, the clear improvement lays in robustness and reproducibility, that was already shown to be far better than with classical systems.

4. Conclusion

The coupling of microfluidic devices with ESI-MS has been pioneered in the early nineties to bring the analytical power of mass spectrometry to previously developed CE microchips. Though most early works were realised in glass microfluidic devices for

proofs-of-principle demonstrations, through years the field crystallised around two kinds of platforms: first, microchips that are devoted to high-throughput sample infusion such as the ESI-Chip from Advion; in this case, many nozzles (at present up to 400 on the ESI-Chip) have to be integrated on the same chip. At the same time, integration of many nozzles requires high manufacturing precision and reproducibility in order to be able to compare nozzle-to-nozzle results. Fabrication of such devices is thus naturally oriented toward silicon-based technologies, that usually have a relatively high cost, but better precisions that polymer based technologies. The other type of devices consists in more integrated ones that embed a liquid-chromatography stage, such as the platform that is being launched by Agilent. Due to the relatively high area of the device, it can hardly be made in silicon at a reasonable cost, and polymer microtechnologies appear as the most viable option from an economic point of view.

5. References

[1] Laurell, T., Marko-Varga, G., Miniaturisation is mandatory unravelling the human proteome, *Proteomics*, **2002**, 2(4): 345-351.

[2] Lion, N., Reymond, F., Girault, H. H., Rossier, J. S., Why the move to microfluidics for protein analysis?, *Curr. Opin. Biotechnol.*, **2004**, 15(1): 31-37.

[3] Saito, Y., Jinno, K., Greibrokk, T., Capillary-columns in liquid chromatography: between conventional columns and microchips, *J. Sep. Sci.*, **2004**, 27(17-18): 1379-1390.

[4] Rozing, G., Nagele, E., Horth, P., Vollmer, M., Moritz, R., Glatz, B., Gratzfeld-Husgen, A., Instrumentation for advanced micro separations in pharmaceutical analysis and proteomics, *J. Biochem. Biophys. Methods*, **2004**, 60(3): 233-263.

[5] Hierlemann, A., Brand, O., Hagleitner, C., Baltes, H., Microfabrication techniques for chemical/biosensors, *Proc. IEEE*, **2003**, 91(6): 839-863.

[6] Voldman, J., Gray, M. L., Schmidt, M. A., Microfabrication in biology and medicine, *Annu. Rev. Biomed. Eng.*, **1999**, 1(401-425.

[7] Kutter, J. P., Current developments in electrophoretic and chromatographic separation methods on microfabricated devices, *Trac-Trends Anal. Chem.*, **2000**, 19(6): 352-363.

[8] Figeys, D., Pinto, D., Proteomics on a chip: Promising developments, *Electrophoresis*, **2001**, 22(2): 208-216.

[9] Reyes, D. R., Iossifidis, D., Auroux, P. A., Manz, A., Micro total analysis systems. 1. Introduction, theory, and technology, *Anal. Chem.*, **2002**, 74(12): 2623-2636.

[10] Auroux, P. A., Iossifidis, D., Reyes, D. R., Manz, A., Micro total analysis systems. 2. Analytical standard operations and applications, *Anal. Chem.*, **2002**, 74(12): 2637-2652.

[11] Lion, N., Rohner, T. C., Dayon, L., Arnaud, I. L., Damoc, E., Youhnovski, N., Wu, Z. Y., Roussel, C., Josserand, J., Jensen, H., Rossier, J. S., Przybylski, M., Girault, H. H., Microfluidic systems in proteomics, *Electrophoresis*, **2003**, 24(21): 3533-3562.

[12] Weigl, B. H., Bardell, R. L., Cabrera, C. R., Lab-on-a-chip for drug development, *Adv. Drug Deliv. Rev.*, **2003**, 55(3): 349-377.

[13] Lee, S. J., Lee, S. Y., Micro total analysis system (mu-TAS) in biotechnology, *Appl. Microbiol. Biotechnol.*, **2004**, 64(3): 289-299.

[14] Barry, R., Ivanov, D., Microfluidics in biotechnology, *J. Nanobiotech.*, **2004**, 2(1): 2.

[15] Vilkner, T., Janazek, D., Manz, A., Micro total analysis systems. Recent developments, *Anal. Chem.*, **2004**, 76(12): 3373-3386.

[16] Livshits, M. A., Ivanov, I. B., Mirzabekov, A. D., Florentev, V. L., DNA-sequencing by hybridization with oligonucleotide matrix (Shom) - Theory of washing out the DNA after hybridization, *Mol. Biol.*, **1992**, 26(6): 856-865.

[17] Fodor, S. P. A., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T., Solas, D., Light-directed, spatially addressable parallel chemical synthesis, *Science*, **1991**, 251(4995): 767-773.

[18] Strezoska, Z., Paunesku, T., Radosavljevic, D., Labat, I., Drmanac, R., Crkvenjakov, R., DNA sequencing by hybridization - 100 bases read by a non-gel- based method, *Proc. Natl. Acad. Sci. U. S. A.*, **1991**, 88(22): 10089-10093.

[19] Manz, A., Fettinger, J. C., Verpoorte, E., Ludi, H., Widmer, H. M., Harrison, D. J., Micromachining of monocrystalline silicon and glass for chemical analysis systems. A look into next century's technology or just a fashionable craze?, *TRAC-Trends in Anal. Chem.*, **1991**, 144-149.

[20] Manz, A., Harrison, D. J., Verpoorte, E. M. J., Fettinger, J. C., Paulus, A., Ludi, H., Widmer, H. M., Planar chips technology for miniaturization and integration of separation techniques into monitoring systems. Capillary electrophoresis on a chip, *J. Chromatogr.*, **1992**, 253-258.

[21] Jacobson, S. C., Hergenroder, R., Koutny, L. B., Ramsey, J. M., Open-channel electrochromatography on a microchip, *Anal. Chem.*, **1994**, 66(14): 2369-2373.

[22] Jacobson, S. C., Hergenroder, R., Koutny, L. B., Ramsey, J. M., High-speed separations on a microchip, *Anal. Chem.*, **1994**, 66(7): 1114-1118.

[23] Woolley, A. T., Mathies, R. A., Ultra-high-speed DNA fragment separations using microfabricated capillary array electrophoresis chips, *Proc. Natl. Acad. Sci. U. S. A.*, **1994**, 91(24): 11348-11352.

[24] Malmqvist, M., Karlsson, R., Biomolecular interaction analysis: affinity biosensor technologies for functional analysis of proteins, *Curr. Opin. Chem. Biol.*, **1997**, 1(3): 378-383.

[25] Martynova, L., Locascio, L. E., Gaitan, M., Kramer, G. W., Christensen, R. G., MacCrehan, W. A., Fabrication of plastic microfluid channels by imprinting methods, *Anal. Chem.*, **1997**, 69(23): 4783-4789.

[26] Bernard, A., Michel, B., Delamarche, E., Micromosaic immunoassays, *Anal. Chem.*, **2001**, 73(1): 8-12.

[27] Hatch, A., Kamholz, A. E., Hawkins, K. R., Munson, M. S., Schilling, E. A., Weigl, B. H., Yager, P., A rapid diffusion immunoassay in a T-sensor, *Nat. Biotechnol.*, **2001**, 19(5): 461-465.

[28] Rossier, J. S., Girault, H. H., Enzyme linked immunosorbent assay on a microchip with electrochemical detection, *Lab Chip*, **2001**, 1(153-157.

[29] Ramsey, R. S., Ramsey, J. M., Generating electrospray from microchip devices using electroosmotic pumping, *Anal. Chem.*, **1997**, 69(6): 1174-1178.

[30] Xue, Q. F., Dunayevskiy, Y. M., Foret, F., Karger, B. L., Integrated multichannel microchip electrospray ionization mass spectrometry: Analysis of peptides from on-chip tryptic digestion of melittin, *Rapid Commun. Mass Spectrom.*, **1997**, 11(12): 1253-1256.

[31] Xue, Q. F., Foret, F., Dunayevskiy, Y. M., Zavracky, P. M., McGruer, N. E., Karger, B. L., Multichannel microchip electrospray mass spectrometry, *Anal. Chem.*, **1997**, 69(3): 426-430.

[32] Oleschuk, R. D., Harrison, D. J., Analytical microdevices for mass spectrometry, *Trac-Trends Anal. Chem.*, **2000**, 19(6): 379-388.

[33] de Mello, A. J., Chip-MS: Coupling the large with the small, *Lab Chip*, **2001**, 1(1): 7N-12N.

[34] Limbach, P. A., Meng, Z. J., Integrating micromachined devices with modern mass spectrometry, *Analyst*, **2002**, 127(6): 693-700.

[35] Figeys, D., Gygi, S. P., McKinnon, G., Aebersold, R., An integrated microfluidics tandem mass spectrometry system for automated protein analysis, *Anal. Chem.*, **1998**, 70(18): 3728-3734.

[36] Figeys, D., Aebersold, R., Nanoflow solvent gradient delivery from a microfabricated device for protein identifications by electrospray ionization mass spectrometry, *Anal. Chem.*, **1998**, 70(18): 3721-7..

[37] Pinto, D. M., Ning, Y. B., Figeys, D., An enhanced microfluidic chip coupled to an electrospray Qstar mass spectrometer for protein identification, *Electrophoresis*, **2000**, 21(1): 181-190.

[38] Meng, Z. J., Qi, S. Z., Soper, S. A., Limbach, P. A., Interfacing a polymer-based micromachined device to a nanoelectrospray ionization Fourier transform ion cyclotron resonance mass spectrometer, *Anal. Chem.*, **2001**, 73(6): 1286-1291.

[39] Li, J. J., Thibault, P., Bings, N. H., Skinner, C. D., Wang, C., Colyer, C., Harrison, J., Integration of microfabricated devices to capillary electrophoresis-electrospray mass spectrometry using a low dead volume connection: Application to rapid analyses of proteolytic digests, *Anal. Chem.*, **1999**, 71(15): 3036-3045.

[40] Vrouwe, E. X., Gysler, J., Tjaden, U. R., van der Greef, J., Chip-based capillary electrophoresis with an electrodeless nanospray interface, *Rapid Commun. Mass Spectrom.*, **2000**, 14(18): 1682-1688.

[41] Xu, N. X., Lin, Y. H., Hofstadler, S. A., Matson, D., Call, C. J., Smith, R. D., A microfabricated dialysis device for sample cleanup in electrospray ionization mass spectrometry, *Anal. Chem.*, **1998**, 70(17): 3553-3556.

[42] Xiang, F., Lin, Y. H., Wen, J., Matson, D. W., Smith, R. D., An integrated microfabricated device for dual microdialysis and on-line ESI ion trap mass spectrometry for analysis of complex biological samples, *Anal. Chem.*, **1999**, 71(8): 1485-1490.

[43] Chartogne, A., Tjaden, U. R., Van der Greef, J., A free-flow electrophoresis chip device for interfacing capillary isoelectric focusing on-line with electrospray mass spectrometry, *Rapid Commun. Mass Spectrom.*, **2000**, 14(14): 1269-1274.

[44] Gao, J., Xu, J. D., Locascio, L. E., Lee, C. S., Integrated microfluidic system enabling protein digestion, peptide separation, and protein identification, *Anal. Chem.*, **2001**, 73(11): 2648-2655.

[45] Wachs, T., Henion, J., Electrospray device for coupling microscale separations and other miniaturized devices with electrospray mass spectrometry, *Anal. Chem.*, **2001**, 73(3): 632-638.

[46] Kameoka, J., Craighead, H. G., Zhang, H. W., Henion, J., A polymeric microfluidic chip for CE/MS determination of small molecules, *Anal. Chem.*, **2001**, 73(9): 1935-1941.

[47] Deng, Y. Z., Henion, J., Li, J. J., Thibault, P., Wang, C., Harrison, D. J., Chip-based capillary electrophoresis/mass spectrometry determination of carnitines in human urine, *Anal. Chem.*, **2001**, 73(3): 639-646.

[48] Deng, Y. Z., Zhang, N. W., Henion, J., Chip-based quantitative capillary electrophoresis/mass spectrometry determination of drugs in human plasma, *Anal. Chem.*, **2001**, 73(7): 1432-1439.

[49] Benetton, S., Kameoka, J., Tan, A. M., Wachs, T., Craighead, H., Henion, J. D., Chip-based P450 drug metabolism coupled to electrospray ionization-mass spectrometry detection, *Anal. Chem.*, **2003**, 75(23): 6430-6436.

[50] Tan, A. M., Benetton, S., Henion, J. D., Chip-based solid-phase extraction pretreatment for direct electrospray mass spectrometry analysis using an array of monolithic columns in a polymeric substrate, *Anal. Chem.*, **2003**, 75(20): 5504-5511.

[51] Lazar, I. M., Ramsey, R. S., Ramsey, J. M., On-chip proteolytic digestion and analysis using "wrong-way-round" electrospray time-of-flight mass spectrometry, *Anal. Chem.*, **2001**, 73(8): 1733-1739.

[52] Chen, S. H., Sung, W. C., Lee, G. B., Lin, Z. Y., Chen, P. W., Liao, P. C., A disposable poly(methylmethacrylate)-based microfluidic module for protein identification by nanoelectrospray ionization- tandem mass spectrometry, *Electrophoresis*, **2001**, 22(18): 3972-3977.

[53] Chiou, C. H., Lee, G. B., Hsu, H. T., Chen, P. W., Liao, P. C., Micro devices integrated with microchannels and electrospray nozzles using PDMS casting techniques, *Sens. Actuator B-Chem.*, **2002**, 86(2-3): 280-286.

[54] Ssenyange, S., Taylor, J., Harrison, D. J., McDermott, M. T., A glassy carbon microfluidic device for electrospray mass spectrometry, *Anal. Chem.*, **2004**, 76(8): 2393-2397.

[55] Lazar, I. M., Ramsey, R. S., Sundberg, S., Ramsey, J. M., Subattomole-sensitivity microchip nanoelectrospray source with time-of-flight mass spectrometry detection, *Anal. Chem.*, **1999**, 71(17): 3627-3631.

[56] Lazar, I. M., Ramsey, R. S., Jacobson, S. C., Foote, R. S., Ramsey, J. M., Novel microfabricated device for electrokinetically induced pressure flow and electrospray ionization mass spectrometry, *J. Chromatogr. A*, **2000**, 892(1-2): 195-201.

[57] Wu, H., Zhai, J., Tian, Y., Lu, H., Wang, X., Jia, W., Liu, B., Yang, P., Xu, Y., Wang, H., Microfluidic enzymatic reactors for peptide mapping: strategy, characterization, and performance, *Lab Chip*, **2004**, 4(588-597.

[58] Sung, W. C., Huang, S. Y., Liao, P. C., Lee, G. B., Li, C. W., Chen, S. H., Poly(dimethylsiloxane)-based microfluidic device with electrospray ionization-mass spectrometry interface for protein identification, *Electrophoresis*, **2003**, 24(21): 3648-3654.

[59] Lazar, I. M., Li, L. J., Yang, Y., Karger, B. L., Microfluidic device for capillary electrochromatography-mass spectrometry, *Electrophoresis*, **2003**, 24(21): 3655-3662.

[60] Li, J. J., Wang, C., Kelly, J. F., Harrison, D. J., Thibault, P., Rapid and sensitive separation of trace level protein digests using microfabricated devices coupled to a quadrupole-time-of-flight mass spectrometer, *Electrophoresis*, **2000**, 21(1): 198-210.

[61] Li, J. J., Kelly, J. F., Chemushevich, I., Harrison, D. J., Thibault, P., Separation and identification of peptides from gel-isolated membrane proteins using a microfabricated device for combined capillary electrophoresis/nanoelectrospray mass spectrometry, *Anal. Chem.*, **2000**, 72(3): 599-609.

[62] Wang, C., Oleschuk, R., Ouchen, F., Li, J. J., Thibault, P., Harrison, D. J., Integration of immobilized trypsin bead beds for protein digestion within a microfluidic chip incorporating capillary electrophoresis separations and an electrospray mass spectrometry interface, *Rapid Commun. Mass Spectrom.*, **2000**, 14(15): 1377-1383.

[63] Li, J. J., Tremblay, T. L., Wang, C., Attiya, S., Harrison, D. J., Thibault, P., Integrated system for high-throughput protein identification using a microfabricated device coupled to capillary electrophoresis/nanoelectrospray mass spectrometry, *Proteomics*, **2001**, 1(8): 975-986.

[64] Li, J. J., LeRiche, T., Tremblay, T. L., Wang, C., Bonneil, E., Harrison, D. J., Thibault, P., Application of microfluidic devices to proteomics research - Identification of trace-level protein digests and affinity capture of target peptides, *Mol. Cell. Proteomics*, **2002**, 1(2): 157-168.

[65] Tachibana, Y., Otsuka, K., Terabe, S., Arai, A., Suzuki, K., Nakamura, S., Robust and simple interface for microchip electrophoresis-mass spectrometry, *J. Chromatogr. A*, **2003**, 1011(1-2): 181-192.

[66] Tachibana, Y., Otsuka, K., Terabe, S., Arai, A., Suzuki, K., Nakamura, S., Effects of the length and modification of the separation channel on microchip electrophoresis-mass spectrometry for analysis of bioactive compounds, *J. Chromatogr. A*, **2004**, 1025(2): 287-296.

[67] Rohner, T. C., Lion, N., Girault, H. H., Electrochemical and theoretical aspects of electrospray ionization, *Phys. Chem. Chem. Phys.*, **2004**, 6(12): 3056-3068.

[68] Pfeifer, R. J., Hendricks, C. D., Parametric studies of electrohydrodynamic spraying, *AIAA J.*, **1968**, 6(3): 496-&.

[69] de la Mora, J. F., The effect of charge emission from electrified liquid cones, *J. Fluid Mech.*, **1992**, 243(561-574.

[70] Emmett, M. R., Caprioli, R. M., Micro-electrospray mass spectrometry - ultra-high-sensitivity analysis of peptides and proteins, *J. Am. Soc. Mass Spectrom.*, **1994**, 5(7): 605-613.

[71] Wilm, M. S., Mann, M., Electrospray and Taylor-cone theory, Doles beam of macromolecules at last, *Int. J. Mass Spectrom. Ion Process.*, **1994**, 136(2-3): 167-180.

[72] New Objective, http://www.newobjective.com.

[73] Tang, K. Q., Lin, Y. H., Matson, D. W., Kim, T., Smith, R. D., Generation of multiple electrosprays using microfabricated emitter arrays for improved mass spectrometric sensitivity, *Anal. Chem.*, **2001**, 73(8): 1658-1663.

[74] Wang, Y. X., Cooper, J. W., Lee, C. S., DeVoe, D. L., Efficient electrospray ionization from polymer microchannels using integrated hydrophobic membranes, *Lab Chip*, **2004**, 4(4): 363-367.

[75] Jackson, G. S., Enke, C. G., Electrical equivalence of electrospray ionization with conducting and nonconducting needles, *Anal. Chem.*, **1999**, 71(17): 3777-3784.

[76] Van Berkel, G. J., Zhou, F. M., Characterization of an electrospray ion-source as a controlled-current electrolytic cell, *Anal. Chem.*, **1995**, 67(17): 2916-2923.

[77] Bard, A. J., Faulkner, L. R., Electrochemical methods: fundamentals and applications, John Wiley & Sons, Inc., 2001.

[78] Schultz, G. A., Corso, T. N., Prosser, S. J., Zhang, S., A fully integrated monolithic microchip electrospray device for mass spectrometry, *Anal. Chem.*, **2000**, 72(17): 4058-4063.

[79] Van Pelt, C. K., Zhang, S., Fung, E., Chu, I. H., Liu, T. T., Li, C., Korfmacher, W. A., Henion, J., A fully automated nanoelectrospray tandem mass spectrometric method for analysis of Caco-2 samples, *Rapid Commun. Mass Spectrom.*, **2003**, 17(14): 1573-1578.

[80] Scholz, M., Gatzek, S., Sterling, A., Fiehn, O., Selbig, J., Metabolite fingerprinting: detecting biological samples by independent component analysis, *Bioinformatics*, **2004**, 20(15): 2447-2454.

[81] Zhang, S., Van Pelt, C. K., Wilson, D. B., Quantitative determination of noncovalent binding interactions using automated nanoelectrospray mass spectrometry, *Anal. Chem.*, **2003**, 75(13): 3010-3018.

[82] Keetch, C. A., Hernandez, H., Sterling, A., Baumert, M., Allen, M. H., Robinson, C. V., Use of a microchip device coupled with mass spectrometry for ligand screening of a multi-protein target, *Anal. Chem.*, **2003**, 75(18): 4937-4941.

[83] Zamfir, A., Vakhrushev, S., Sterling, A., Niebel, H. J., Allen, M., Peter-Katalinic, J., Fully automated chip-based mass spectrometry for complex carbohydrate system analysis, *Anal. Chem.*, **2004**, 76(7): 2046-2054.

[84] Bindila, L., Almeida, R., Sterling, A., Allen, M., Peter-Katalinic, J., Zamfir, A., Off-line capillary electrophoresis/fully automated nanoelectrospray chip quadrupole time-of-flight mass spectrometry and tandem mass spectrometry for glycoconjugate analysis, *J. Mass Spectrom.*, **2004**, 39(1190-1201.

[85] Zamfir, A., Vukelic, Z., Bindila, L., Peter-Katalinic, J., Almeida, R., Sterling, A., Allen, M., Fully-automated chip-based nanoelectrospray tandem mass spectrometry of gangliosides from human cerebellum, *J. Am. Soc. Mass Spectrom.*, **2004**, 15(1649-1657.

[86] Zhang, S., Van Pelt, C. K., Henion, J. D., Automated chip-based nanoelectrospray-mass spectrometry for rapid identification of proteins separated by two-dimensional gel electrophoresis, *Electrophoresis*, **2003**, 24(21): 3620-3632.

[87] Zhang, S., Chelius, D., Characterization of protein glycosylation using chip-based infusion nanoelectrospray linear ion trap tandem mass spectrometry, *J. Biomol. Techn.*, **2004**, 15(2): 120-33.

[88] Leuthold, L. A., Grivet, C., Allen, M., Baumert, M., Hopfgartner, G., Simultaneous selected reaction monitoring, MS/MS and MS3 quantitation for the analysis of pharmaceutical compounds in human plasma using chip-based infusion, *Rapid Commun. Mass Spectrom.*, **2004**, 18(17): 1995-2000.

[89] Kapron, J. T., Pace, E., Van Pelt, C. K., Henion, J., Quantitation of midazolam in human plasma by automated chip-based infusion nanoelectrospray tandem mass spectrometry, *Rapid Commun. Mass Spectrom.*, **2003**, 17(18): 2019-2026.

[90] Dethy, J. M., Ackermann, B. L., Delatour, C., Henion, J. D., Schultz, G. A., Demonstration of direct bioanalysis of drugs in plasma using nanoelectrospray infusion from a silicon chip coupled with tandem mass spectrometry, *Anal. Chem.*, **2003**, 75(4): 805-811.

[91] Chen, J., Yang, L., Kapron, J. T., Ma, L., Pace, E., Van Pelt, C. K., Rudewicz, P. J., Determination of SCH 211803 by nanoelectrospray infusion mass spectrometry:

evaluation of matrix effect and comparison with liquid-chromatography-tandem mass spectrometry, *J. Chromatogr. B*, **2004**, 809(205-210.

[92] Froesch, M., Bindila, L. M., Baykut, G., Allen, M., Peter-Katalinic, J., Zamfir, A. D., Coupling of fully automated chip electrospray to Fourier transform ion cyclotron resonance mass spectrometry for high-performance glycoscreening and sequencing, *Rapid Commun. Mass Spectrom.*, **2004**, 18(3084-3092.

[93] Staack, R. F., Varesio, E., Hopfgartner, G., The combination of liquid chromatography/tandem mass spectrometry and chip-based infusion for improved screening and characterization of drug metabolites, *Rapid Commun. Mass Spectrom.*, **2005**, 19(618-626.

[94] Gabelica, V., Vreuls, C., Filee, P., Duval, V., Joris, B., De Pauw, E., Advantages and drawbacks of nanospray for studying noncovalent protein-DNA complexes by mass spectrometry, *Rapid Commun. Mass Spectrom.*, **2002**, 16(18): 1723-1728.

[95] Benkestock, K., Van Pelt, C. K., Akerud, T., Sterling, A., Edlund, P. O., Roeraade, J., Automated nano-electrospray mass spectrometry for protein-ligand screening by noncovalent interaction applied to human H-FABP and A-FABP, *J. Biomol. Screen*, **2003**, 8(3): 247-256.

[96] De Vriendt, K., Sandra, K., Desmet, T., Nerinckx, W., Van Beeumen, J., Devreese, B., Evaluation of automated nano-electrospray mass spectrometry in the determination of non-covalent protein-ligand complexes, *Rapid Commun. Mass Spectrom.*, **2004**, 18(24): 3061-3067.

[97] Wickremsinhe, E. R., Ackermann, B. L., Chaudhary, A. K., Validating regulatory-compliant wide dynamic range bioanalytical assays using chip-based nanoelectrospray tandem mass spectrometry, *Rapid Commun. Mass Spectrom.*, **2005**, 19(47-56.

[98] Kalkum, M., Lyon, G. J., Chait, B. T., Detection of secreted peptides by using hypothesis-driven multistage mass spectrometry, *Proc. Natl. Acad. Sci. U. S. A.*, **2003**, 100(5): 2795-2800.

[99] Griss, P., Melin, J., Sjodahl, J., Roeraade, J., Stemme, G., Development of micromachined hollow tips for protein analysis based on nanoelectrospray ionization mass spectrometry, *J. Micromech. Microeng.*, **2002**, 12(5): 682-687.

[100] Sjodahl, J., Melin, J., Griss, P., Emmer, A., Stemme, G., Roeraade, J., Characterization of micromachined hollow tips for two-dimensional nanoelectrospray mass spectrometry, *Rapid Commun. Mass Spectrom.*, **2003**, 17(4): 337-341.

[101] Ramsey, R. S., Ramsey, J. M., Generating electrospray from microchip devices using electroosmotic pumping (vol 69, pg 1174, 1997), *Anal. Chem.*, **1997**, 69(13): 2617-2617.

[102] Licklider, L., Wang, X. Q., Desai, A., Tai, Y. C., Lee, T. D., A micromachined chip-based electrospray source for mass spectrometry, *Anal. Chem.*, **2000**, 72(2): 367-375.

[103] Kim, J. S., Knapp, D. R., Miniaturized multichannel electrospray ionization emitters on poly(dimethylsiloxane) microfluidic devices, *Electrophoresis*, **2001**, 22(18): 3993-3999.

[104] Kim, J. S., Knapp, D. R., Microfabrication of polydimethylsiloxane electrospray ionization emitters, *J. Chromatogr. A*, **2001**, 924(1-2): 137-145.

[105] Kim, J. S., Knapp, D. R., Microfabricated PDMS multichannel emitter for electrospray ionization mass spectrometry, *J. Am. Soc. Mass Spectrom.*, **2001**, 12(4): 463-469.

[106] Yuan, C. H., Shiea, J., Sequential electrospray analysis using sharp-tip channels fabricated on a plastic chip, *Anal. Chem.*, **2001**, 73(6): 1080-1083.
[107] Rohner, T. C., Rossier, J. S., Girault, H. H., Polymer Microspray with an Integrated Thick-Film Microelectrode, *Anal. Chem.*, **2001**, 73(22): 5353-5357.
[108] Gobry, V., van Oostrum, J., Martinelli, M., Rohner, T. C., Reymond, F., Rossier, J. S., Girault, H. H., Microfabricated polymer injector for direct mass spectrometry coupling, *Proteomics*, **2002**, 2(4): 405-412.
[109] Muck, A., Svatos, A., Atmospheric molded poly(methylmethacrylate) microchip emitters for sheathless electrospray, *Rapid Commun. Mass Spectrom.*, **2004**, 18(13): 1459-1464.
[110] do Lago, C. L., da Silva, H. D. T., Neves, C. A., Brito-Neto, J. G. A., da Silva, J. A. F., A dry process for production of microfluidic devices based on the lamination of laser-printed polyester films, *Anal. Chem.*, **2003**, 75(15): 3853-3858.
[111] Kameoka, J., Orth, R., Ilic, B., Czaplewski, D., Wachs, T., Craighead, H. G., An electrospray ionization source for integration with microfluidics, *Anal. Chem.*, **2002**, 74(22): 5897-5901.
[112] Yang, Y. N., Kameoka, J., Wachs, T., Henion, J. D., Craighead, H. G., Quantitative mass spectrometric determination of methylphenidate concentration in urine using an electrospray ionization source integrated with a polymer microchip, *Anal. Chem.*, **2004**, 76(9): 2568-2574.
[113] Lion, N., Gellon, J. O., Girault, H. H., Flow-rate characterization of microfabricated polymer microspray emitters, *Rapid Commun. Mass Spectrom.*, **2004**, 18(14): 1614-1620.
[114] Rossier, J. S., Youhnovski, N., Lion, N., Damoc, E., Reymond, F., Girault, H. H., Przybylski, M., Thin-chip microspray system for coupling with high performance Fourier-Transform Ion Cyclotron Resonance mass spectrometry, *Ang. Chem. Intl. Ed.*, **2003**, 42(1): 53-58.
[115] Bindila, L., Froesch, M., Lion, N., Vukelic, Z., Rossier, J. S., Girault, H. H., Peter-Katalinic, J., Zamfir, A., Thin chip microsprayer system coupled to Fourier transform ion cyclotron resonance mass spectrometry for glycoconjugate screening, *Rapid Comm. Mas Spectrom.*, **2004**, in press.
[116] Zamfir, A., Lion, N., Vukelic, Z., Bindila, L., Rossier, J. S., Girault, H. H., Peter-Katalinic, J., Thin chip microsprayer system coupled to quadrupole time-of-flight mass spectrometer for glycoconjugate analysis, *Lab Chip*, **2005**, 5(3): 298-307.
[117] Lion, N., Gobry, V., Jensen, H., Rossier, J. S., Girault, H., Integration of a membrane-based desalting step in a microfabricated disposable polymer injector for mass spectrometric protein analysis, *Electrophoresis*, **2002**, 23(20): 3583-3588.
[118] Lion, N., Gellon, J. O., Jensen, H., Girault, H. H., On-chip protein sample desalting and preparation for direct coupling with electrospray ionization mass spectrometry, *J. Chromatogr. A*, **2003**, 1003(1-2): 11-19.
[119] Rohner, T. C., Rossier, J. S., Girault, H. H., On-line electrochemical tagging of cysteines in proteins during nanospray, *Electrochem. Commun.*, **2002**, 4(9): 695-700.
[120] Roussel, C., Rohner, T. C., Jensen, H., Girault, H. H., Mechanistic aspects of on-line electrochemical tagging of free L-cysteine residues during electrospray ionisation for mass spectrometry in protein analysis, *ChemPhysChem*, **2003**, 4(2): 200-206.
[121] Rohner, T. C., Josserand, J., Jensen, H., Girault, H. H., Numerical investigation of an electrochemically induced tagging in a nanospray for protein analysis, *Anal. Chem.*, **2003**, 75(9): 2065-2074.

[122] Roussel, C., Dayon, L., Jensen, H., Girault, H. H., On-line cysteine modification for protein analysis: new probes for electrochemical tagging nanospray mass spectrometry, *J. Electroanal. Chem.*, **2004**, 570(2): 187-199.
[123] Dayon, L., Roussel, C., Prudent, M., Lion, N., Girault, H. H., On-line counting of cysteine residues in peptides during electrospray ionisation by electrogenerated tags and its application to protein identification, *Electrophoresis*, **2005**, 26(1): 238-247.
[124] Roussel, C., Dayon, L., Lion, N., Rohner, T. C., Josserand, J., Rossier, J. S., Jensen, H., Girault, H. H., Generation of mass-tags by the inherent electrochemistry of electrospray for protein mass spectrometry, *J. Amer. Soc. Mass Spectrom.*, **2004**, 15(12): 1767-1779.
[125] Wen, J., Lin, Y. H., Xiang, F., Matson, D. W., Udseth, H. R., Smith, R. D., Microfabricated isoelectric focusing device for direct electrospray ionization-mass spectrometry, *Electrophoresis*, **2000**, 21(1): 191-197.
[126] Svedberg, M., Pettersson, A., Nilsson, S., Bergquist, J., Nyholm, L., Nikolajeff, F., Markides, K., Sheathless electrospray from polymer microchips, *Anal. Chem.*, **2003**, 75(15): 3934-3940.
[127] Svedberg, M., Veszelei, M., Axelsson, J., Vangbo, M., Nikolajeff, F., Poly(dimethylsiloxane) microchip: microchannel with integrated open electrospray tip, *Lab Chip*, **2004**, 4(4): 322-327.
[128] Le Gac, S., Arscott, S., Cren-Olive, C., Rolando, C., Two-dimensional microfabricated sources for nanoelectrospray, *J. Mass Spectrom.*, **2003**, 38(12): 1259-1264.
[129] Le Gac, S., Arscott, S., Rolando, C., A planar microfabricated nanoelectrospray emitter tip based on a capillary slot, *Electrophoresis*, **2003**, 24(21): 3640-3647.
[130] Arscott, S., Le Gac, S., Druon, C., Tabourier, P., Rolando, C., Micromachined 2D nanoelectrospray emitter, *Electron. Lett.*, **2003**, 39(24): 1702-1703.
[131] Le Gac, S., Cren-Olive, C., Rolando, C., Arscott, S., A novel nib-like design for microfabricated nanospray tips, *J. Am. Soc. Mass Spectrom.*, **2004**, 15(3): 409-412.
[132] Arscott, S., Le Gac, S., Druon, C., Tabourier, P., Rolando, C., A planar on-chip micro-nib interface for NanoESI-MS microfluidic applications, *J. Micromech. Microeng.*, **2004**, 14(2): 310-316.
[133] Arscott, S., Le Gac, S., Druon, C., Tabourier, P., Rolando, C., A micro-nib nanoelectro spray source for mass spectrometry, *Sens. Actuator B-Chem.*, **2004**, 98(2-3): 140-147.
[134] Le Gac, S., Rolando, C., Arscott, S., Novel 2D ionization sources for nano electrospray mass spectrometry (ESI-MS), in *Proc. 5th International conference on thermal and mechanical simulation and experiments in microelectronics and microsystems EuroSimE2004*, 2004.
[135] Brinkmann, M., Blossey, R., Arscott, S., Druon, C., Tabourier, P., Le Gac, S., Rolando, C., Microfluidic design rules for capillary slot-based electrospray sources, *Appl. Phys. Lett.*, **2004**, 85(11): 2140-2142.
[136] Yin, N. F., Killeen, K., Brennen, R., Sobek, D., Werlich, M., van de Goor, T. V., Microfluidic chip for peptide analysis with an integrated HPLC column, sample enrichment column, and nanoelectrospray tip, *Anal. Chem.*, **2005**, 77(2): 527-533.
[137] Rozing, G., van de Goor, T., Yin, H. F., Killeen, K., Glatz, B., Kraiczek, K., Lauer, H. H., An experimental study of chromatographic dynamics in open and packed non-cylindrical conduits, *J. Se.p Sci.*, **2004**, 27(17-18): 1391-1401.
[138] Fortier, M. H., Bonneil, E., Goodley, P., Thibault, P., Integrated microfluidic device for mass spectrometry-based proteomics and its application to biomarker discovery programs, *Anal. Chem.*, **2005**, 77(6): 1631-1640.

[139] Ninonuevo, M., Yin, H., An, H., Killeen, K., Grimm, R., Ward, R., German, B., Lebrilla, C., Nano-liquid chromatography/mass spectrometry of oligosaccharides employing porous graphitized carbon column on microchip with high accuracy mass analyzer, *Electrophoresis*, **2005**, in press.

Chapter 3. Coupling of microfluidic devices with electrospray mass spectrometers

1. Introduction

In this chapter, polymer microfluidic chips are validated as microsprayers for the analysis of peptides and proteins by ESI-MS. In the first part, chip microsprayers are used as stand-alone devices: a sample reservoir is pasted on the chip to allow sample loading, which generates a gravity-induced flow within the microchannel. The chip has been used in this configuration to analyse peptides and proteins by Fourier Transform-Ion Cyclotron Resonance mass spectrometry through a collaboration with Pr Przybylski's group (University of Konstanz, Germany) [1]. Alternatively, in the second part chips are coupled to infusion lines in order to better control the sample flow rate [2], or ultimately be coupled to liquid-phase separations to serve as interfaces for LC-MS or CE-MS.

2. Microchips used as stand-alone microsprayers

2.1. Material and methods

Microchip fabrication

Microchips used in this work were kindly provided by DiagnoSwiss (Monthey, Switzerland). They consist in polyimide foils in which microchannels (120 µm wide, 45 µm

deep, 1 cm long with a half moon cross section, see Figure 26 below) are ablated. Five gold-coated microelectrodes sit at the bottom of each microchannel. The microchip fabrication process has been described in [3, 4]. Briefly, the process starts with 30 x 40 cm polyimide foils which are coated on each side by a 5 µm copper layer.

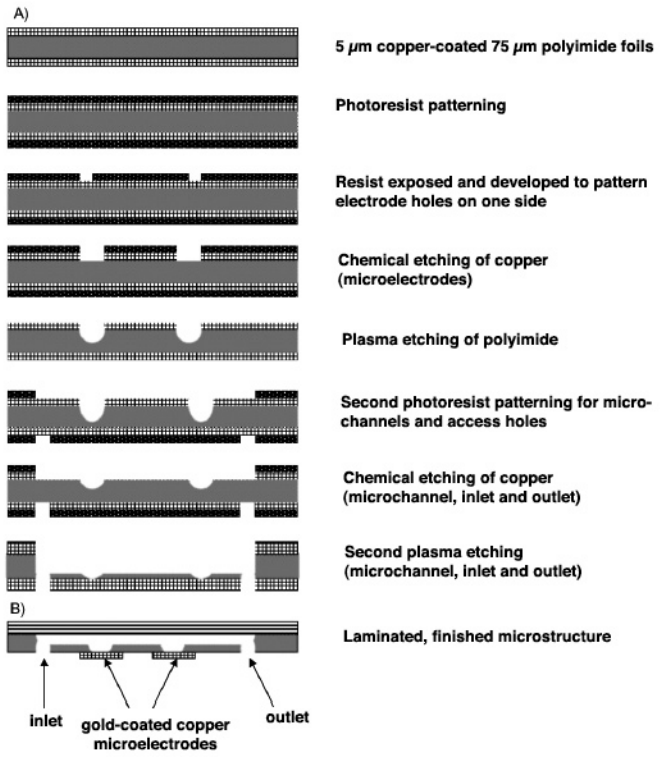


Figure 25. Scheme of the fabrication process of polyimide microchips by plasma etching.

A photoresist is patterned, and exposed and developed to expose copper where microelectrodes and access holes will be etched. The copper layer is then chemically etched with an aqueous solution of $CuCl_2$ and H_2O_2. Polyimide is then exposed to a low pressure plasma formed in an O_2/CF_4 gas mixture with an O_2 content of about 80%. Isotropic polyimide etching occurs at an ablation rate of about 1-2 $\mu m.min^{-1}$. A second photoresist is patterned, exposed and developed to expose copper where microchannels and access holes will be etched. A second stage of chemical etching (to remove the exposed copper) and plasma etching (to ablate polyimide) results in the structure depicted in the last row of

Figure 25A. The remaining polyimide over the microelectrode is laser ablated to expose copper, on which gold is subsequently electroplated. The resulting gold-coated microelectrodes are 50 µm in diameter. Finally, microchips are laminated with a 35 µm thick polyethylene/ polyethylene terephthalate foil with a standard lamination device (Morane, Oxon, UK), the top roll being heated at 130°C and the bottom one at 52°C, the lamination layer being rolled over the microchip at a pressure of 2 bars.

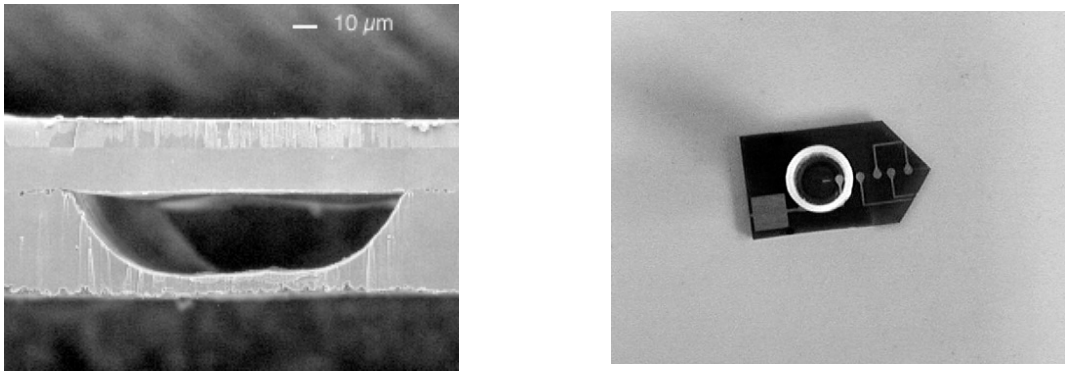


Figure 26. Left: Scanning Electron Microscopy of a polymer microchip. Right: Polymer microspray emitter showing the tip shape outlet, the gold conductive tracks to the five embedded microelectrodes, and the pasted sample reservoir.

In order to open one end of the microchannel, one extremity of the microchip is cut in a tip shape, the microchannel outlet being at the apex of the tip, as previously described [5, 6]. In order to load the microchip, a polystyrene reservoir is pasted over the microchannel inlet: the reservoir is first dipped into chloroform to melt its bottom and then manually applied on the microchip surface till it is pasted. The reservoir itself has an inner diameter of 7 mm.

Samples and reagents

Oxidised bovine insulin B-chain, human angiotensin I and bovine ubiquitin were obtained from Sigma (Saint Louis, MO, USA) and sequencing grade modified trypsin from Promega (Madison, WI, USA). Single peptides and peptide libraries comprising the mucin-

2 epitope motif, TQTXT, were synthesized by solid-phase peptide synthesis (SPPS) on a chloromethylated polystyrene resin with the portioning-mixing method, using Boc/benzyl (Boc1/tert-butoxycarbonyl) protection chemistry [7]. Coupling steps were monitored by ninhydrin and the Boc group removed with 33% trifluoroacetic acid (TFA). Final cleavage from the resin was performed by treatment with anhydrous HF for 1 h, and crude libraries and single peptides were purified by HPLC before analysis by MALDI-TOF-MS and ESI-FT-ICR-MS.

Mass spectrometry

The chip-ESI-FT-ICR-MS coupling interface was constructed in a self-built chip holder, prepared for fixation of up to five chips. The holder is equipped with x,y,z-fine adjustment screws for movement and positioning of the chips in front of the capillary entrance. The ESI voltage was applied onto the chip contacts or by using a platinum electrode plunged in the chip reservoir. The standard entrance of the ESI ion source was removed, and placed 1-2 mm in front of the capillary entrance.

Mass spectra were obtained with a 7T Bruker Daltonik Apex II FTICR mass spectrometer equipped with an APOLLO electrospray nano-ESI source. Sample solutions of 0.5-5 μL (0.0005-0.01 $\mu g.mL^{-1}$) were applied to the chip reservoir, and the chip fixed into the sample holder. The spray voltage was slowly increased from 0 V to the working voltage of 1000-2300 V, and the onset of the spray inspected with a microscope. Mass spectra were obtained by acquisition of typically, 4-32 single scans at 45-70 V capillary exit voltage, 9.14 V of skimmer 1, 8.82 V of skimmer 2, 2500 ms ionization pulse time, 0.001 s ionization delay time, 1.2 ms excitation sweep pulse. Oxidized bovine insulin B-chain was dissolved in water/methanol/acetic acid (49.5/49.5/1) to a final concentration of 0.0071 $\mu g.mL^{-1}$. Bovine ubiquitin was dissolved in a solution of 3% aqueous acetic acid/methanol

(1/1) to yield a final concentration of 0.002 μg.mL^{-1}. In-gel digestion with trypsin (2 h) and sample preparation of human tau protein was performed as previously described [8-10]. Infrared multiphoton dissociation (IRMPD) experiments were performed with Ar as the cooling gas at the following experimental steps: 5000 ms empty ESI source, empty ICR cell: 5000 ms hexapole accumulation; 1800 ms ion injection, 40000 ms IR laser in cell; 30 ms detection delay. The collision-induced dissociation (CID) experiments were performed at a capillary exit voltage of 70±120 V.

2.2. Results

Fluid mobilisation

Fluid mobilisation in polymer microfluidic channels depend on electroosmotic flow, and externally imposed flow. As well known, electroosmotic flow is highly dependent on pH:

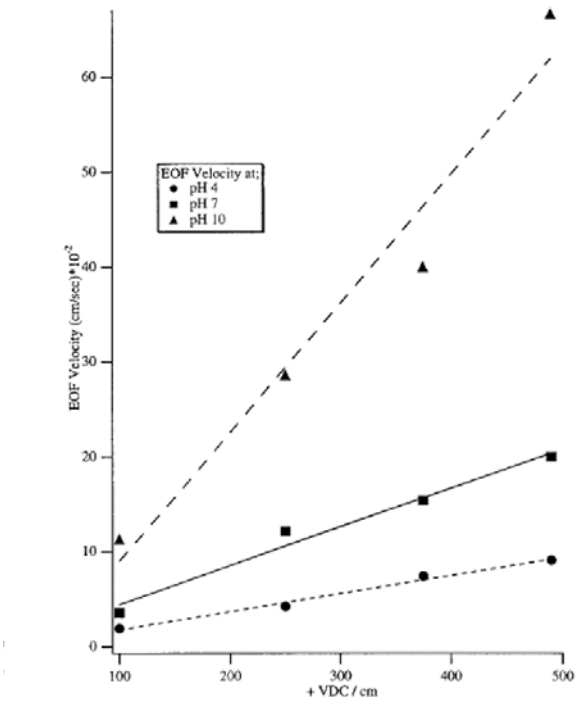


Figure 27. Dependency of the electroosmotic velocity vs applied electric field curves on pH. Reprinted with permission from [11].

With the microchannel sizes used in this study, these values correspond to flow rates in the range 10 nL/min (pH 4 at 100 V/cm) up to 2 µL/min (pH 10, 500 V/cm). In most cases, the solutions used in this chapter are acidified to pH 2.6 (1% acetic acid), which means that there is virtually no electroosmotic flow present.

Flow rates within the microchannels can also be induced by the height of liquid present in the reservoir. The height of liquid is directly related to the volume of sample loaded:

$$H = \frac{V}{\pi \frac{d^2}{4}} \quad (1)$$

Where V is the volume of sample loaded, d the reservoir diameter and H the height of liquid within the reservoir. Pascal's law relates the pressure induced by the height of liquid:

$$\Delta P = \rho g H \quad (2)$$

Where ρ is the fluid density ($kg.m^{-3}$), g is the gravity acceleration ($m.s^{-2}$). Finally Poiseuille equation relates the flow rate induced by a given pressure ΔP to the geometrical parameters:

$$Q = \frac{\Delta P \pi R^4}{8 \eta l} \quad (3)$$

Where Q is the induced flow rate, R the microchannel radius, η the fluid viscosity (Pa.s), and l the microchannel length. Combining equations (1)-(3) gives the relationship between the sample volume loaded into the reservoir and the flow rate in the microchannel:

$$Q = \frac{R^4 \rho g}{2 \eta l d^2} V \quad (4)$$

The validity of equation (4) has been validated experimentally by Rossier *et al* [4, 12]. Physical parameters for methanol/water mixtures are given Table 2:

	Density (×1000 $kg.m^{-3}$)	Viscosity (Pa.s)	$\frac{R^4 \rho g}{2 \eta l d^2} (s^{-1})$
Water at 20°C	0.998	1.00×10^{-3}	1.29×10^{-4}
Methanol at 20°C	0.792	5.84×10^{-4}	1.76×10^{-4}
Methanol/water (50:50)	0.895	7.92×10^{-4}	1.47×10^{-4}

Table 2. Physical parameters for methanol/water solutions.

With the given geometry, gravity-induced flow rates within the microchannel are shown in Figure 28:

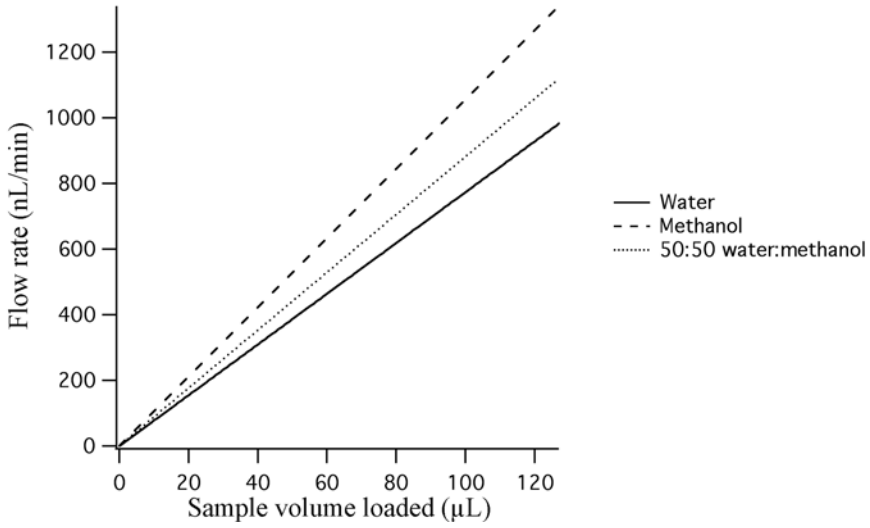


Figure 28. Gravity-induced flow rates versus sample volume loaded in the reservoir.

Unless otherwise stated, sample volumes loaded into the reservoir are of 20 µL, which correspond to flow rates of 155 nL/min (pure water), 211 nL/min (pure methanol) and 176 nL/min (50:50 methanol:water mixture).

Peptide mass spectrometry

Examples of chip-ESI-FT-ICR spectra are shown in Figure 29 for bovine insulin B-chain and ubiquitin. The chip FT-ICR-MS measurements provided mass determination accuracies of 1.4 ppm for insulin and 2.0 ppm for ubiquitin by using routine broad band (BB) mode and side-kick (SK) trapping method, at mass resolutions in the BB mode of 57000 (ubiquitin) and 91000 (insulin B-chain).

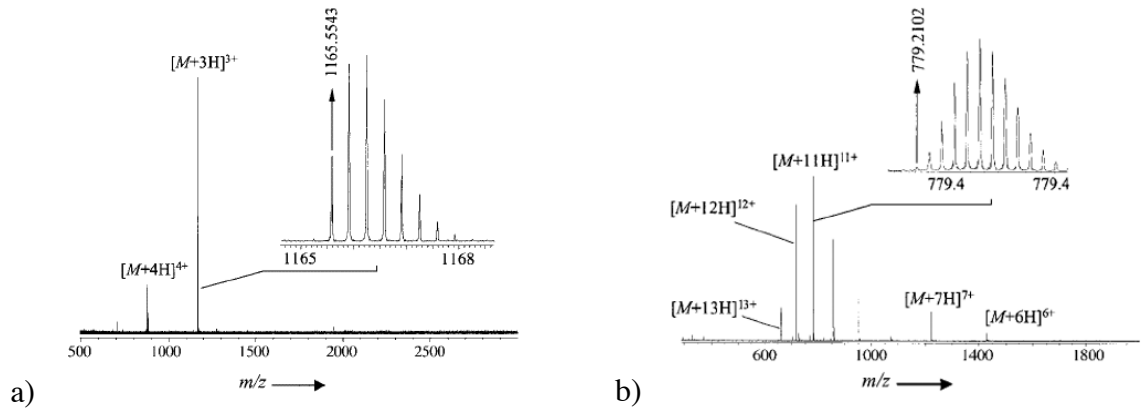


Figure 29. Chip-ESI-FT-ICR mass spectra of a) oxidized bovine insulin B chain, (7.1 µg/mL in 2% aqueous acetic acid/methanol (1/1), b) bovine ubiquitin (20. µg/mL in 3% aqueous acetic acid/methanol (1/1).

However, mass resolution is readily increased up to $>10^6$ by using the high-resolution mode. At standard resolution (BB mode) the charge states of ions were readily defined from the mass difference of two adjacent isotopes, without the necessity of deconvolution techniques. This situation is of importance, for example, for the analysis of noncovalent complexes where a low charge distribution may yield only few peaks that are difficult to deconvolute. The high resolution of the chip-FT-ICR spectra routinely provides monoisotopic peaks with mass-determination accuracies around 1 ppm. The characterization of complex mixtures, such as from natural products and combinatorial libraries, often requires preliminary fractionation or separation (e.g. by HPLC or CE). In such applications the high resolution and accuracy of FT-ICR-MS proves to be extremely useful, for example, to detect mixture components with closely adjacent masses [13, 14]. The feasibility of the chip-ESI-FT-ICR system is illustrated in Figure 30 for a combinatorial peptide library derived from the core epitope of mucin-2 gastrointestinal glycoprotein, a marker protein for colon cancer [7].

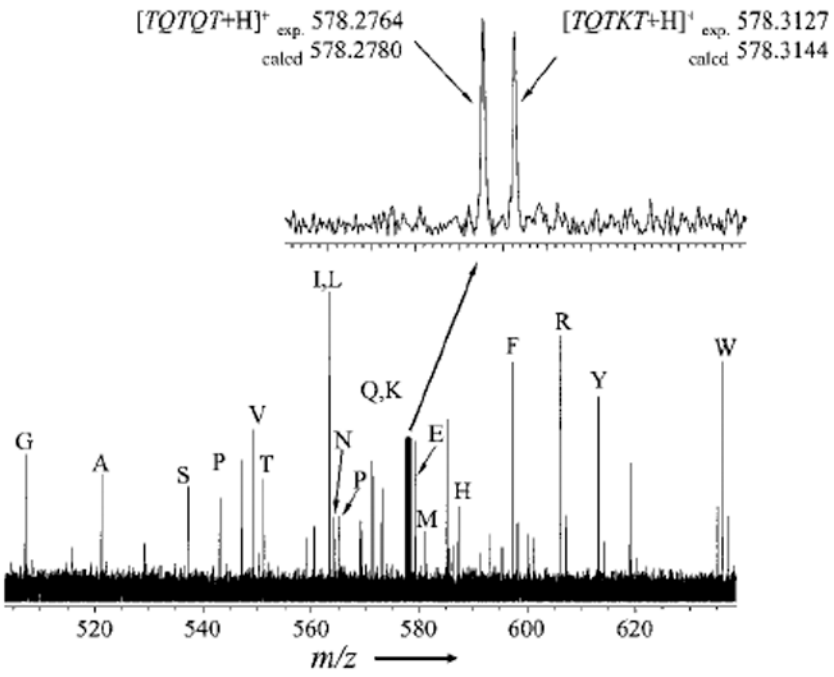


Figure 30. Chip-ESI-FT-ICR mass spectrum of the combinatorial peptide sub-library TQTXT (where X = all proteinogenic amino acids except Cys). The peaks corresponding to the isobaric peptides TQTIT and TQTLT overlap. The insert shows the base-line separated peaks corresponding to the protonated molecular ions of peptides TQTQT and TQTKT (see Table 3).

In a combinatorial peptide sub-library of the epitope motif TQTXT, the chip-ESI-FTICR-MS analysis was capable of distinguishing, and provided identification of all components of the library at monoisotopic resolution. Protonated molecular ions were observed with a relative-abundance distribution reflecting the composition of components, except for the peptide TQTHT, whereas the ion signal of the two isobaric peptides, TQTIT and TQTLT, were twice as abundant as the other components. The chip-ESI spectrum provided unequivocal resolution of the "isobaric" peptides, TQTKT and TQTQT with experimental molecular masses, 578,31271 and 578.27636 (Δm=0.03635 amu). The accuracies of mass determinations yielded the elemental compositions of all the library components (Table 3).

Amino acid X	$[TQTXT+H]^+_{calc}$	$[TQTXT+H]^+_{exp}$	Δm	Mass accuracy (ppm)	Elemental composition
A	521.25712	521.25547	0.00165	3.1	$C_{20}H_{37}N_6O_{10}$
D	565.24695	565.24558	0.00137	2.4	$C_{21}H_{36}N_6O_{12}$
E	579.26260	579.26044	0.00216	3.7	$C_{22}H_{39}N_6O_{12}$
F	597.28842	597.28656	0.00186	3.1	$C_{26}H_{41}N_6O_{10}$
G	507.24147	507.23953	0.00194	3.8	$C_{19}H_{35}N_6O_{10}$
H	587.27891	587.27640	0.00251	4.3	$C_{23}H_{39}N_8O_{10}$
I	563.30407	563.30182	0.00225	4.0	$C_{23}H_{43}N_6O_{10}$
K	578.31497	578.31271	0.00226	3.9	$C_{23}H_{44}N_7O_{10}$
L	563.30407	563.30182	0.00225	4.0	$C_{23}H_{43}N_6O_{10}$
M	581.26049	581.25877	0.00172	3.0	$C_{22}H_{41}N_6O_{10}S_1$
N	564.26293	564.26118	0.00175	3.1	$C_{21}H_{38}N_7O_{11}$
P	547.27277	547.27080	0.00197	3.6	$C_{22}H_{39}N_6O_{10}$
Q	578.27858	578.27636	0.00222	3.8	$C_{22}H_{40}N_7O_{11}$
R	606.32111	606.31884	0.00227	3.7	$C_{23}H_{44}N_9O_{10}$
S	537.25203	537.24984	0.00219	4.1	$C_{20}H_{37}N_6O_{11}$
T	551.26768	551.26583	0.00185	3.4	$C_{21}H_{39}N_6O_{11}$
V	549.28842	549.28628	0.00204	3.7	$C_{22}H_{41}N_6O_{10}$
W	636.29932	636.29657	0.00275	4.3	$C_{28}H_{42}N_7O_{10}$
Y	613.28333	613.28049	0.00284	4.6	$C_{26}H_{41}N_6O_{11}$
		Average	0.00210	3.7	
		Standard deviation	0.0004	0.5	

Table 3. Elemental compositions of components of the TQTXT library determined by chip ESI-FT-ICR-MS.

Similar results were obtained for a number of polypeptides and proteins within an approximate molecular-weight range between 1 and 35 kDa (data not shown). The chip ESI-FT-ICR-MS interface provides a stable spray of 20 µL sample solutions for at least 15 min, which enables analysis of low sample concentrations, that is, attomole sensitivity for oligopeptides as shown previously for nano-ESI-FT-ICR-MS [15-17]. However, for peptides in the lower mass range a significant advantage of the chip-ESI-system was found in the absence of the background and/or contamination that is frequently found with nano-ESI capillaries. Furthermore, spraying times with the chip are sufficient to carry out the complete range of fragmentation experiments, one of the key features of FT-ICRMS [18-25]. Fragmentation experiments of the model peptide angiotensin were carried out in the

ESI source low vacuum region (source CID), as well as in the ICR-cell, after isolation of ions using 1) Ar as collision gas (SORI-CID), or 2) IRMPD. The MS/MS feasibility of the chip-ESIFT- ICR system was demonstrated by IRMPD fragmentation of angiotensin yielding results comparable to those obtained using the nano-ESI source.

MS/MS fragmentation	Isolated ions	Sequence determined	Resolution	Accuracy (ppm)
IRMPD	$[M+2H]^{2+}$ $[M+3H]^{3+}$	VYIHP	50000	6
Source CID	—	IHPFHL	60000	3

Table 4. Fragmentation of angiotensin I (DRVYIHPFHL) by in-source SORI-CID and in-cell IRMPD.

First multistage (MS^n) experiments were performed up to the MS^3 level by using both SORI-CID and IRMPD. A first application of the chip- FTICR system to a complex multicomponent mixture in proteome analysis is shown in Figure 31 with an in-gel digested tryptic peptide mixture of the human microtubule associated tau protein which has recently aroused interest as a target protein in neurofibrillary tangles characteristic of Alzheimer disease. The pathophysiological aggregation of tau in fibrillary tangles has been shown to be associated with structural changes by multi-(hyper)- phosphorylation which has therefore become the subject of recent studies.

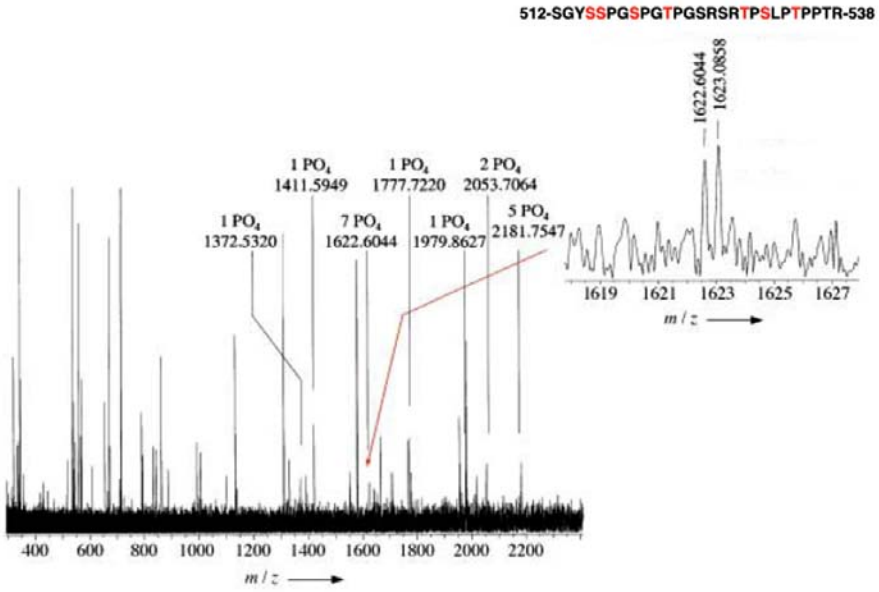


Figure 31. Chip-ESI-FTICR mass spectrum of tryptic digest mixture of multiphosphorylated human neurofibrillary tau protein. Mass assignments with m/z detection range, 200-2500, and partial sequences are shown for phosphorylated peptides. The insert shows the peptide fragment of the phosphorylation domain (512-538); the seven phosphorylation sites, identified using information from the NiceProt View of Swiss-Prot database (primary accession number P10636), are indicated in red ($C_{113}H_{191}N_{36}O_{61}P_7$, doubly protonated; calculated: 1622.55513; found: 1622.60435, Δ=30 ppm).

In the chip-ESI spectrum the complete primary structure including 18 serine- and threonine-phosphorylations could be directly identified from the tryptic peptides, partially with very low abundances, which clearly demonstrates the high analytical performance of the chip-FTICR-MS. Mass determinations yielded accuracies of 5-20 ppm within the phosphorylated peptide sequences; this provided the direct identification of 7 and 5 phosphorylation sites in the multi-phosphorylation domains, (512–538; see insert in Figure 31, and (382–398; T386, S388, T395, T396, S397).

2.3. Conclusion

The goal of this first study was to validate the use of stand-alone microfluidic microsprayer on an ESI source different from the one that had been used in our laboratory, namely the heated capillary ion source of the LCQ Duo (Thermo Electron, San Jose, USA) [5, 6]. The coupling with the Appolo source of the FT-ICR-MS instrument (Bremen, Germany) appeared to be relatively easy, though it required the removal of the entrance cap of the heated capillary and that the microchip positioning in front of the heated capillary inlet proved to be more critical than on the LCQ Duo.

The second validation consisted in the analysis of more complex mixtures (the TQTXT library) than previously analysed, and of real life samples (Tau protein digest from 2D-GE). In both cases, no reversible adsorption of analytes could be observed: even extensive washing of the microchip did not revealed elution of potentially adsorbed analytes (data not shown).

The microfluidic microsprayer used as a stand-alone device appears to be particularly promising when combined with high-resolution FT-ICR-MS, where analytes are delivered to the mass spectrometer without prior separation, and where relatively long analysis times are needed to perform complex MS acquisitions or MS/MS fragmentation processes. In this instance, the low flow rates generated by the height of liquid present in the sample reservoir (below 200 nL/min) combined with stable ion currents shows great promise.

3. Microfluidic microsprayers under infusion

3.1. Materials and methods

Chemicals and reagents

Analytical grade methanol (MeOH) and trifluoroacetic acid (TFA) are from Riedel de Haen (Darmstaadt, Germany), acetic acid and propranolol-S-hydrochloride are from Fluka (Buchs, Switzerland), and HPLC grade acetonitrile (ACN) is from SDS (Peypin, France). Deionised water (18.2 MW.cm) is obtained from a MilliQ unit (Millipore, Bedford, USA).

Microchip fabrication

Microchips were kindly provided by DiagnoSwiss (Monthey, Switzerland). They are exactly the same as the one used in the previous section. Instead of pasting a reservoir on top of the microchip for sample loading, the microchip was sandwiched into a home-built polycarbonate holder that integrates a low dead-volume microfluidic connection to couple the microchip to a syringe pump (KD Scientific, New Hope, PA, USA). The whole device is shown in Figure 32.

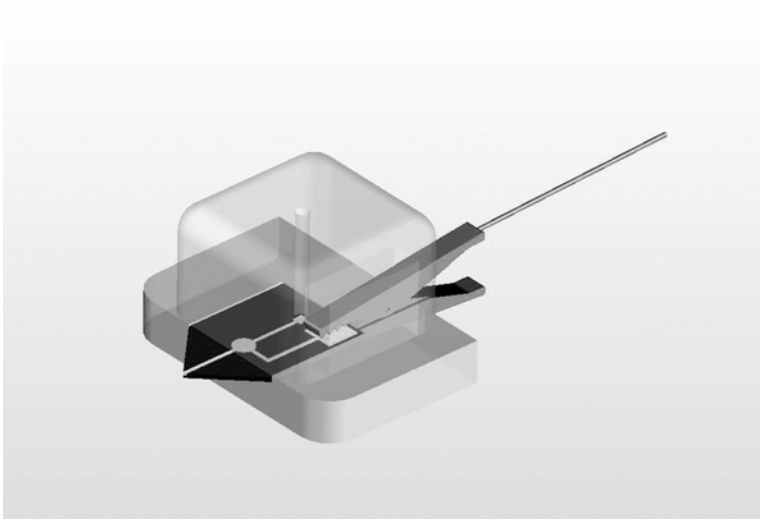


Figure 32. Scheme of the chip holder used to connect an infusion line with the microspray.

Mass spectrometry

Mass spectrometer is a LCQ Duo ion trap from Thermo Electron (San Jose, CA, USA). All experiments were done in full scan mode (50-2000 *m/z*) without any averaging, the heated capillary is kept at 200 °C, and ion optics parameters are optimized on the studied compound. In the polymer microspray experiments, the chip holder is placed in front of the heated capillary inlet, typically 3-5 mm away. The electrospray voltage is applied through the integrated microelectrode, and typically set between 2 and 3 kV. For the pulled nanospray experiments, borosilicate nanospray capillaries (50 µm inner diameter, 3 to 5 µm outlet diameter, from Polymicro, Phoenix, AZ, USA) are mounted on the commercial nanospray source, connected to the syringe pump, and electrospray voltage is applied through a liquid junction on the infusion line, and typically set between 1 and 2.5 kV. In each experiment, the flow rate is varied with all other parameters unchanged (chip or nanospray position, electrospray voltage). Experiments were done whether in 50% MeOH: 49% water: 1% acetic acid, or in X% ACN: (100-X-1)% water: 1% TFA.

3.2. Results and discussion

For planar microsprays that do not have a nozzle-like outlet, the hydrophobicity of the microspray material plays a significant role in its ability to accommodate a stable Taylor cone: if the spray solution can wet easily the outer microspray walls, the Taylor cone will spill over on the edge of the microchip, and potentially delocalize away from the microspray outlet, as shown in Figure 33: liquid coming out from the microchannel spilled over on the edge of the chip till it reached an asperity, where the Taylor cone stabilised.

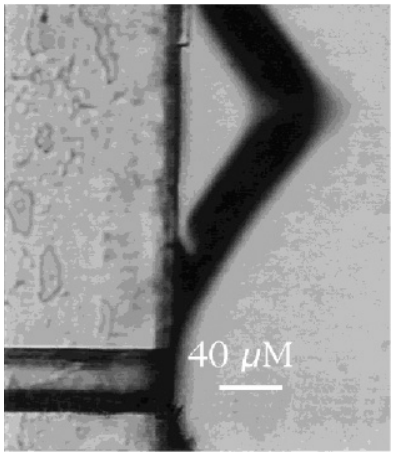


Figure 33. Delocalisation of the Taylor cone on the edge of a PET microspray. Reprinted with permission from [5].

On the contrary, if the spray solution cannot wet the microspray material, the Taylor cone will be confined at the outlet of the microchannel, and its size will vary with the flow rate applied, but it will not collapse easily. Figure 34 shows the contact angles measured for water and two classical spray medium (acidified methanol: water and acidified acetonitrile: water): glass is easily wet by water, methanol and acetonitrile spray media; whereas some authors did not find any improvement in the spray stability with hydrophobic treatments of the outlet of the glass microspray [26], some others found it compulsory for good operation [27, 28].

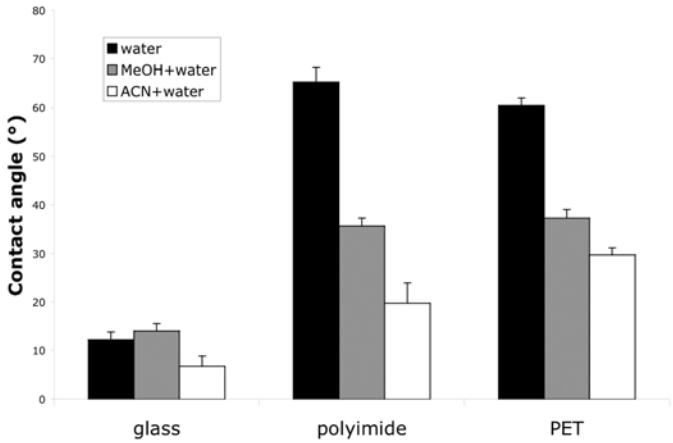


Figure 34. Contact angles of water (black bars), 50% MeOH: 1% acetic acid: water (grey bars), and 50% ACN: 1% TFA (white bars) on bare glass, polyimide (outer surface of the microspray), and polyethylene terephthalate (outer surface of microsprays comparable to those used in our laboratory [5, 29-33]). Error bars represent one standard deviation over triplicate measurements.

Moreover, most glass microfluidic devices incorporate a pasted nanospray tip or are connected to a microelectrospray needle for easy operation [34-37], which may indicate that spraying directly from a glass microchip is not easy. On the contrary, both polyimide and polyethylene terephthalate (PET) show high contact angles for water and decent contact angles for the two spray solutions tested, which makes then particularly suitable materials for polymer microsprays.

The first goal of this study was to explore the flow rate range that can be applied on the microchip. To characterize the flow rate dependency of spectral signal, a solution of 10 µM propranolol in 50% MeOH: 49% water: 1% acetic acid (v:v) is infused at different flow rates. The propranolol concentration was chosen to have high signals without any saturation of the MS detector (saturation in the chosen conditions would be around 10^8 ion counts on our instrument under the chosen conditions). Figure 35 shows the evolution of propranolol ion intensity (measured as peak area at m/z = 260.1) when increasing the flow rate of the infused solution.

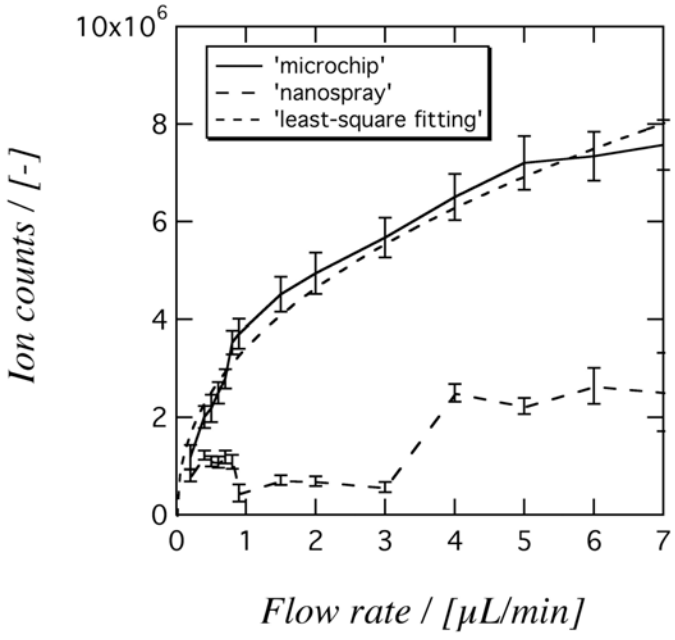


Figure 35. Evolution of the propranolol ion count (m/z = 260.1) with increasing flow rates. Error bars represent one standard deviation of the propranolol peak area over 10 min.

The electrospray signal obtained from the microchip increases smoothly from 250 nL/min up to 7 µL/min while the signal generated from the nanospray capillary becomes more erratic, and never reaches the value obtained from the microchip. Above 7 µL/min, it was impossible to generate electrospray from the nanospray capillary, while spraying was still possible from the microchip with compromised stability; in fact at higher flow rate, a droplet slowly formed at the outlet of the microchannel till electrospray became impossible to maintain. The flow-rate dependency of the propranolol intensity is found to be of the form $I=A.V_f^x$ where V_f is the flow rate, with A=3.45.106±1.17.105 ion counts / µL/min and x=0.43±0.02 (see Figure 35). This dependency is in good agreement with the empirical expression for the electrospray current proposed by Pfeifer *et al* [38]:

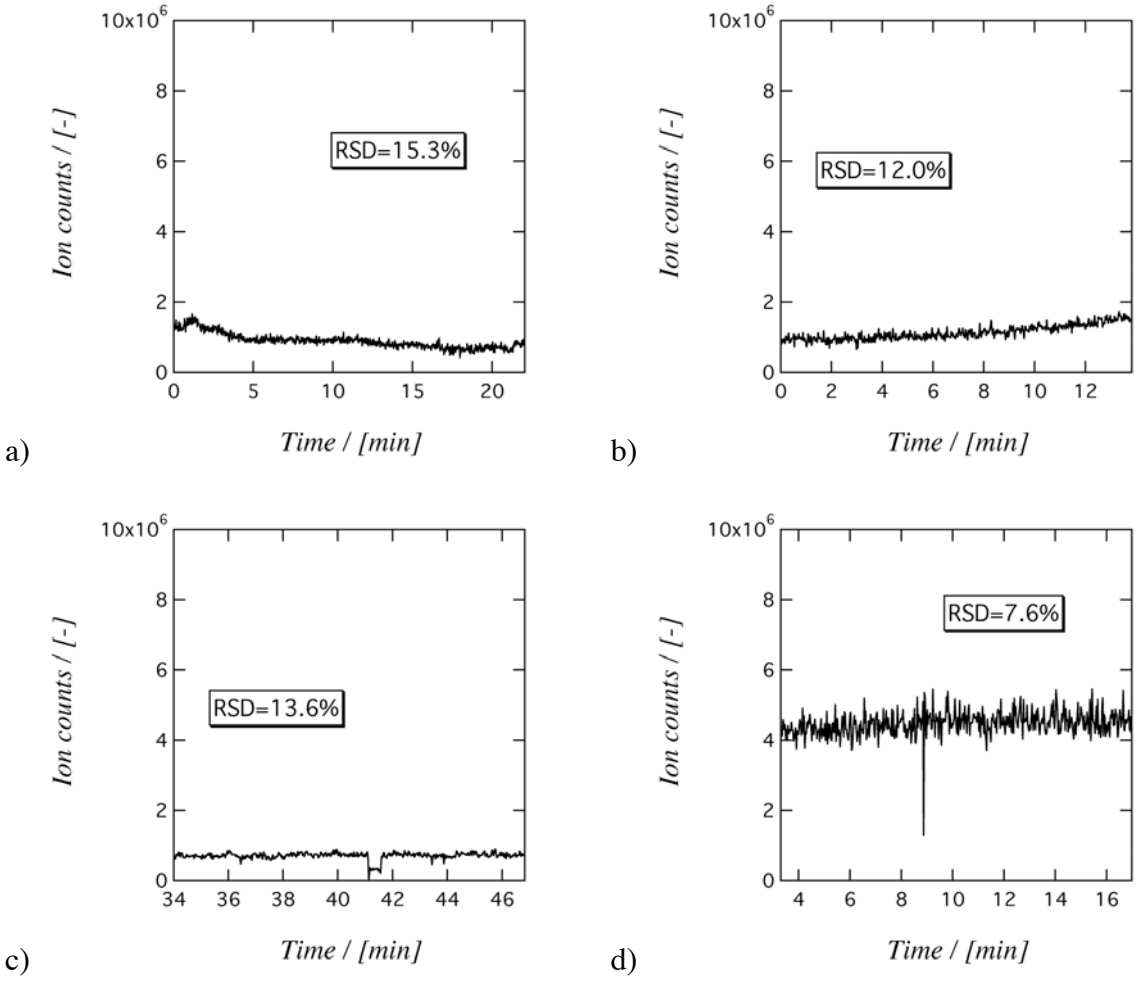
$$i_{es} = \left[\left(\frac{4\pi}{\varepsilon}\right)^3 (9\gamma)^2 \varepsilon_0{}^5\right]^{1/7} (\kappa E)^{3/7} V_f^{4/7} \qquad (5)$$

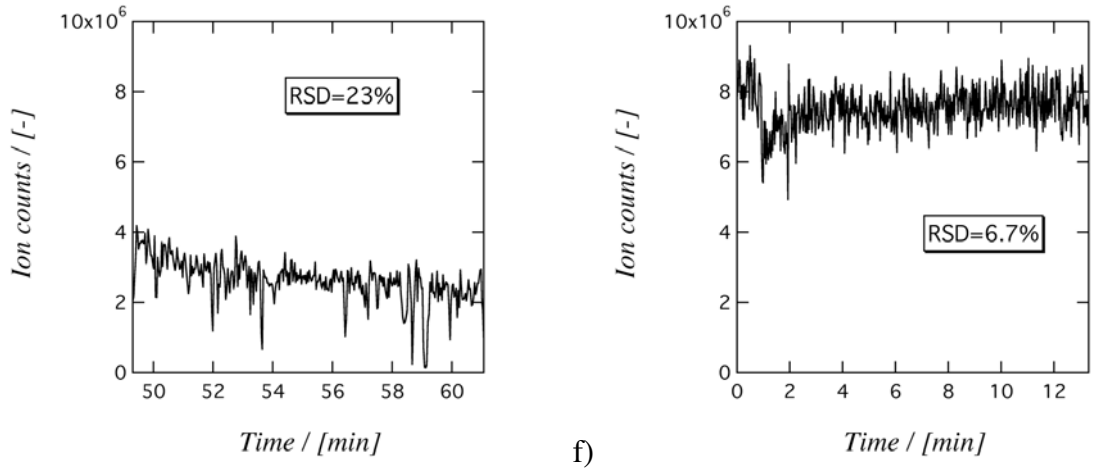
where ε and ε_0 are the permittivity of the solvent and of the vacuum respectively, γ is the surface tension of the solvent, κ is the conductivity of the infused solution and V_f is the volumic flow rate, as well as with the simplified relationship derived by de la Mora *et al* [39]:

$$i_{es} = f(\varepsilon_r).(\gamma \kappa V_f \varepsilon_r)^{0.5} \qquad (6)$$

On the contrary, no global dependency can be outlined for the nanospray capillary because of its instability; this difference in behaviour can probably be attributed to the hydrophobicity of the polymer surface as well as to the shape of the microchip outlet: a big Taylor cone can indeed expand on the microchip outlet walls without spilling over, while the sharp edges of the nanospray capillary cannot accommodate large Taylor cones. Whereas obviously only the lower flow rates are relevant for use with the nanospray capillary, the results are given for the whole flow rate range.

Qualitatively, the signal generated from the microchip was found to be more stable through time than the one generated from the nanospray capillary, especially at high flow rates, as illustrated by Figure 36.





e) f)

Figure 36. Propranolol ion counts (m/z @ 260.1) for signals generated from the nanospray capillary at 0.2 (a), 1.5 (c) and 7 μL/min (e), and from the microchip at 0.2 (b), 1.5 (d) and 7 μL/min (f). Round square deviations (RSD) are estimated over the more stable 10 minutes in all cases.

In order to quantitatively compare the signal stabilities when increasing the flow rate, the temporal signal to noise ratio was calculated as:

$$\mathrm{SNR} = 20.\log\frac{\bar{I}(\text{propanolol})}{\sigma(\text{propanolol})} \qquad \text{in dB} \qquad (7)$$

The mean propranolol intensity and standard deviation were calculated from peak area over 10 minutes in each case. As can be seen from Figure 37, the temporal signal-to-noise ratio shows a 10 dB increase from 0.2 μL/min to 1 μL/min for the signal generated from the microchip, and remains more or less constant for higher flow rates.

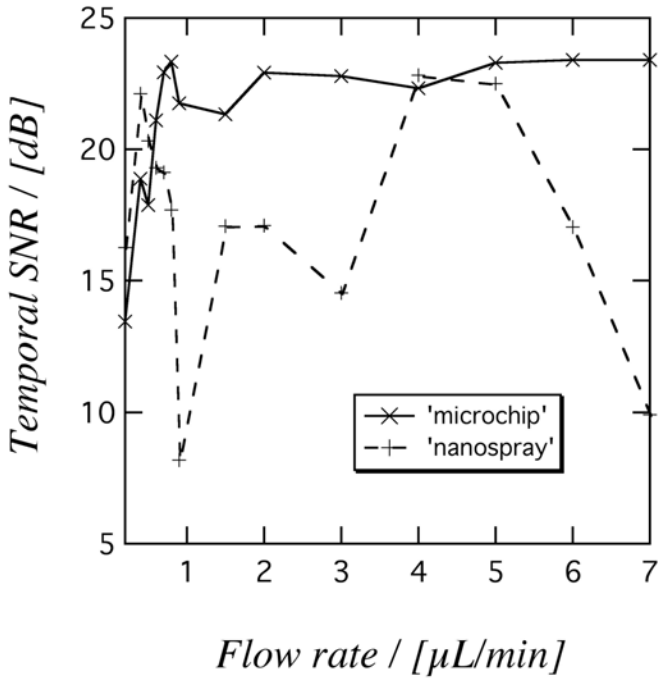


Figure 37. Evolution of the temporal signal-to-noise ratio with increasing flow rates.

On the contrary, the signal generated from the nanospray shows the same increase in temporal signal-to-noise ratio up to 0.7 µL/min but becomes really irregular at higher flow rates. The typical ranges where the microchip and nanospray show best stability are above 1 µL/min and below 0.7 µL/min respectively.

Furthermore, the limit of detection (LOD) was estimated as:

$$\mathrm{LOD} = \overline{\mathrm{noise}} + 3\sigma_{noise} \quad \text{in dB} \qquad (8)$$

where noise mean and noise standard deviation were calculated between m/z = 500 and m/z = 2000 in each spectrum (no analyte peak present). Figure 38 shows the absolute limits of detection for the microchip and the nanospray.

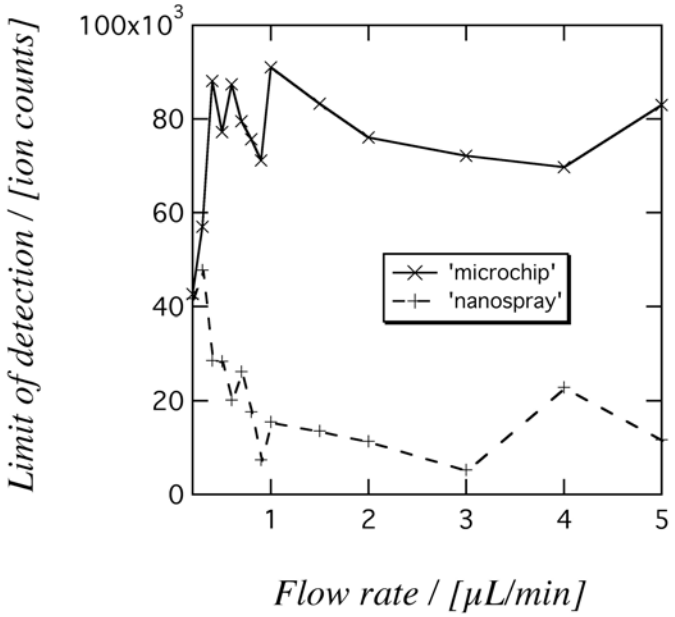


Figure 38. Theoretical limits of detection (LOD) of the microchip and nanospray, in ion counts.

Not surprisingly, the nanospray shows much lower limit of detection at each flow rate (by an approximate factor of four). But a more important feature for mass spectral characterization of the micro- or nanospray emitters is the spectral signal-to-noise ratio, which is estimated as:

$$\mathrm{SNR} = 20.\log\frac{I(\mathrm{propanolol})}{I(noise)} \qquad \text{in dB} \qquad (9)$$

where the propranolol intensity is the peak area at m/z = 260.1, and noise mean intensity calculated as above. This spectral SNR basically reflects the ratio between the analyte signal and spectral noise. Figure 39 shows the evolution of the spectral signal-to-noise ratio for both microchip and nanospray.

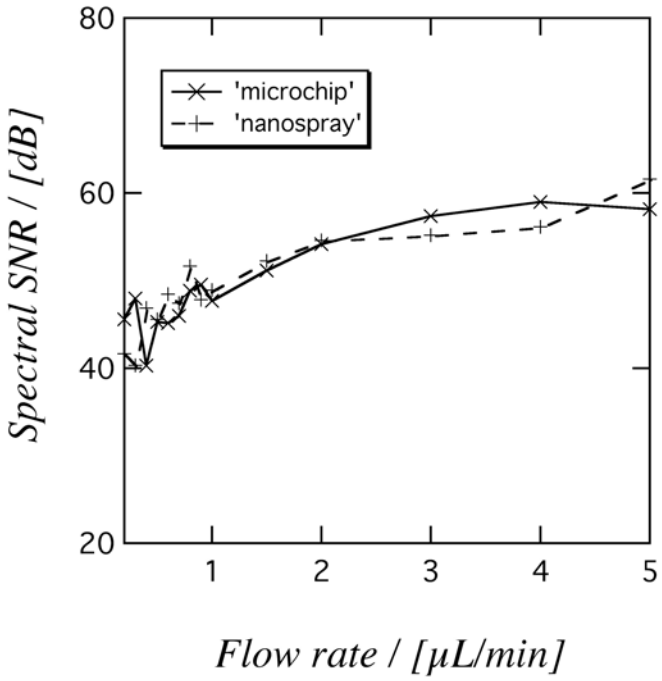


Figure 39. Evolution of the spectral signal-to-noise ratio with increasing flow rates.

Remarkably, the microchip and the nanospray behave pretty much in the same way, except at the highest flow rates tested, where the nanospray shows better SNR (but a poor temporal stability in this domain, see above). Surprisingly, the higher the flow rate, the better the spectral signal-to-noise ratio.

3.3. Compatibility with acetonitrile

In order to investigate the compatibility of the microspray emitter with acetonitrile-containing spraying solution, the same kind of experiments was performed and resulted in comparable results: signals from propranolol in 50% ACN: 49% water: 1% TFA were generated at flow rates between 0.250 and 4 µL/min, as shown in Figure 8. Propranolol intensity increases smoothly up to 4 µL/min; at this point, the sprayed solution began to wet the outer walls of the microchip, so that a small drop slowly formed before being sprayed, resulting in a kind of pulse-mode spray (data not showed).

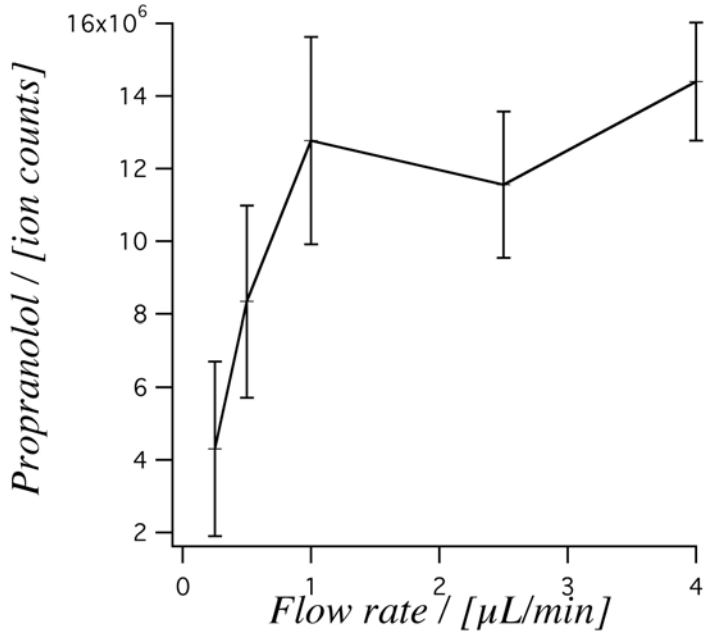


Figure 40. Evolution of the propranolol intensity with increasing flow rates in 50% ACN: 49% water: 1% TFA. Error bars represent one standard deviation of the peak area over 10 min.

This difference of behaviour between methanol and acetonitrile can be attributed to differences in hydrophobicity of the two solvents: acetonitrile wets the hydrophobic polyimide walls more easily than methanol (see Figure 34). The spectral signal-to-noise ratio also increases smoothly with flow rate (see Figure 41), with optimal values above 1 µL/min, as is the case with methanol.

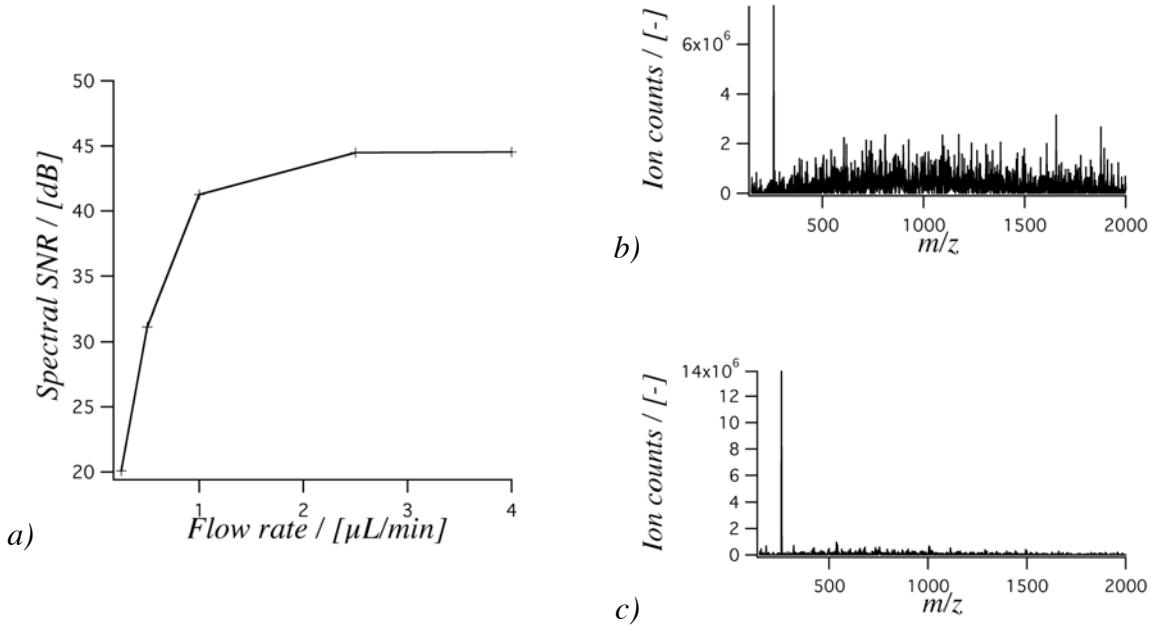


Figure 41. a) Spectral signal-to-noise ratio of propranolol spectra versus flow rate. b) and c) are corresponding spectra at 0.250 and 4 µL/min respectively

Moreover, acetonitrile can be used in high amounts, as shown in Figure 42: the single scan of 10 µM propranolol in 99% ACN: 1% TFA shows the single peak of propranolol and no contaminant from the chip itself even if the spray is operated for 20 minutes, resulting in a spectral SNR of 41.2 dB.

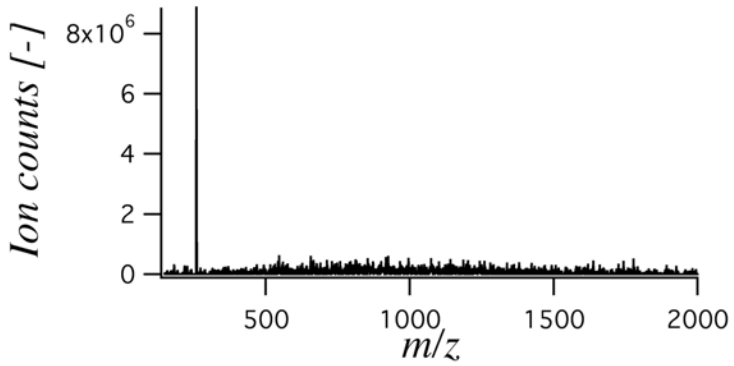


Figure 42. Single spectrum of 10 µM propranolol in 99% ACN: 1% TFA infused at 2.5 µL/min.

On the opposite, satisfactory spray can be obtained even in 10% ACN: 89% water: 1% TFA, as shown in Figure 43.

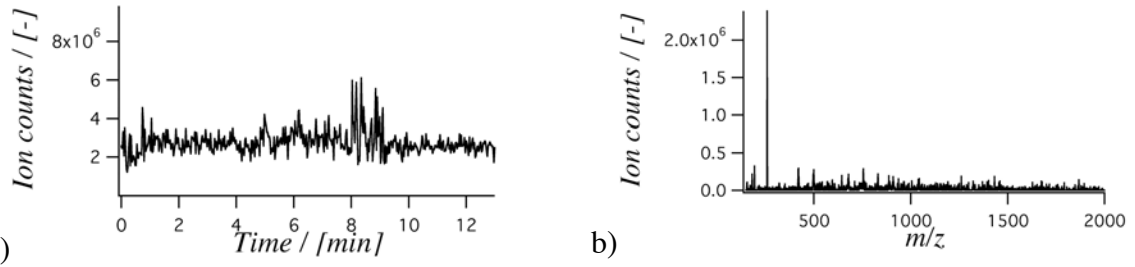


Figure 43. Microchip performances in 10% ACN: 89% water: 1% TFA. a) Stability of the spray over 12 minutes, showing a temporal SNR of 13 dB. b) Single scan showing a spectral SNR of 38.5 dB.

A temporal SNR of 13 dB is obtained for an infusion at 2.5 µL/min, corresponding to the lowest value obtained with 50% MeOH: 49% water: 1% acetic acid over the whole flow rate range (see Figure 37). Nevertheless, the spray is stable and results in a spectral SNR of 38.5 dB, corresponding again to the lowest value obtained with methanol solutions (see Figure 39). Below 10% acetonitrile, it was impossible to spray continuously. The microchip can thus be used in a wide range of solvent compositions (from 10% to 99% of acetonitrile).

3.4. Conclusion

In this section the flow rate behaviour of the microfluidic microsprayer was tested. Despite its relatively large dimensions (120×45 µm) compared to classical pulled nanospray capillaries that have outlet diameters of 3-10 µm typically, the microsprayer shows good sensitivities in the low flow rate range. Additionally, it can accommodate much larger flow rates (up to 7 µL/min) that cannot be used with pulled nanospray capillaries. In summary, the microsprayer shows equivalent performances to pulled nanospray capillaries in terms of

signal stability and sensitivity, and better performances at higher flow rates. It can also be used in a wide range of solvent conditions, from 10% to 99% of organic solvent. It thus appears as a particularly promising interface for LC-MS hyphenation: the flow rate ranges tested here is fully compatible with nano- and capillary LC systems, the range of solvent composition corresponds to gradient conditions of LC separation of peptides and proteins in reverse phase liquid chromatography. From a practical point of view, microfluidic microsprayers have the advantages of robustness and disposability compared to pulled nanospray capillaries. Additionally its larger dimensions prevent any clogging problems that are often encountered with small nanosprays.

4. Bibliography

[1] Rossier, J. S., Youhnovski, N., Lion, N., Damoc, E., Reymond, F., Girault, H. H., Przybylski, M., Thin-chip microspray system for coupling with high performance Fourier-Transform Ion Cyclotron Resonance mass spectrometry, *Ang. Chem. Intl. Ed.*, **2003**, 42(1): 53-58.

[2] Lion, N., Gellon, J. O., Girault, H. H., Flow rate characterization of microfabricated microspray emitters, *Rapid Comm. Mass Spectrom.*, **2004**, 18(14): 1614-1620.

[3] Rossier, J. S., Vollet, C., Carnal, A., Lagger, G., Gobry, V., Girault, H. H., Michel, P., Reymond, F., Plasma etched polymer microelectrochemical systems, *Lab Chip*, **2002**, 2(3): 145-150.

[4] Rossier, J. S., Reymond, F., Michel, P., Polymer microfluidic chips for electrochemical and biochemical analyses, *Electrophoresis*, **2002**, 23(6): 858-867.

[5] Rohner, T. C., Rossier, J. S., Girault, H. H., Polymer Microspray with an Integrated Thick-Film Microelectrode, *Anal. Chem.*, **2001**, 73(22): 5353-5357.

[6] Gobry, V., van Oostrum, J., Martinelli, M., Rohner, T., Rossier, J. S., Girault, H. H., Microfabricated polymer injector for direct mass spectrometry coupling, *Proteomics*, **2002**, 2(4): 405-412.

[7] Uray, K., Price, M. R., Hudecz, F., Localisation of a protein core-specific epitope from gastrointestinal mucin (MUC2). The effect of epitope immobilisation on antibody recognition, *J. Pept. Sci.*, **1998**, 4(5): 319-326.

[8] Meri, S., Baumann, M., Proteomics: posttranslational modifications, immune responses and current analytical tools, *Biomol. Eng.*, **2001**, 18(5): 213-220.

[9] Bauer, S. H. J., Wiechers, M. F., Bruns, K., Przybylski, M., Stuermer, C. A. O., Isolation and identification of the plasma membrane-associated intracellular protein reggie-2 from goldfish brain by chromatography and Fourier-transform ion cyclotron resonance mass spectrometry, *Anal. Biochem.*, **2001**, 298(1): 25-31.

[10] Larsen, M. R., Sorensen, G. L., Fey, S. J., Larsen, P. M., Roepstorff, P., Phospho-proteomics: Evaluation of the use of enzymatic de-phosphorylation and differential

mass spectrometric peptide mass mapping for site specific phosphorylation assignment in proteins separated by gel electrophoresis, *Proteomics*, **2001**, 1(2): 223-238.

[11] Roberts, M. A., Rossier, J. S., Bercier, P., Girault, H., UV laser machined polymer substrates for the development of microdiagnostic systems, *Anal. Chem.*, **1997**, 69(11): 2035-2042.

[12] Morier, P., Vollet, C., Michel, P. E., Reymond, F., Rossier, J. S., Gravity-induced convective flow in microfluidic systems: electrochemical characterisation and application to enzyme-linked immunosorbent assay tests, *Electrophoresis*, **2004**, 25(11): 3761-3768.

[13] Windberg, E., Hudecz, F., Marquardt, A., Sebestyen, F., Kiss, A., Bosze, S., Medzihradszky-Schweiger, H., Przybylski, M., Characterisation of combinatorial libraries of mucin-2 antigen peptides by high-resolution mass spectrometry, *Rapid Commun. Mass Spectrom.*, **2002**, 16(9): 834-839.

[14] Walk, T. B., Trautwein, A. W., Richter, H., Jung, G., ESI fourier transform ion cyclotron resonance mass spectrometry (ESI-FT-ICR-MS): A rapid high-resolution analytical method for combinatorial compound libraries, *Angew. Chem.-Int. Edit.*, **1999**, 38(12): 1763-1765.

[15] Gucek, M., Gaspari, M., Walhagen, K., Vreeken, R. J., Verheij, E. R., van der Greef, J., Capillary electrochromatography/nanoelectrospray mass spectrometry for attomole characterization of peptides, *Rapid Commun. Mass Spectrom.*, **2000**, 14(16): 1448-1454.

[16] Neubauer, G., Mann, M., Mapping of phosphorylation sites of gel-isolated proteins by nanoelectrospray tandem mass spectrometry: Potentials and limitations, *Anal. Chem.*, **1999**, 71(1): 235-242.

[17] Gatlin, C. L., Kleemann, G. R., Hays, L. G., Link, A. J., Yates, J. R., Protein identification at the low femtomole level from silver-stained gels using a new fritless electrospray interface for liquid chromatography microspray and nanospray mass spectrometry, *Anal. Biochem.*, **1998**, 263(1): 93-101.

[18] Marshall, A. G., Hendrickson, C. L., Jackson, G. S., Fourier transform ion cyclotron resonance mass spectrometry: A primer, *Mass Spectrom. Rev.*, **1998**, 17(1): 1-35.

[19] Chalmers, M. J., Hakansson, K., Johnson, R., Smith, R., Shen, J. W., Emmett, M. R., Marshall, A. G., Protein kinase A phosphorylation characterized by tandem Fourier transform ion cyclotron resonance mass spectrometry, *Proteomics*, **2004**, 4(4): 970-981.

[20] Hakansson, K., Emmett, M. R., Marshall, A. G., Davidsson, P., Nilsson, C. L., Structural analysis of 2 D-gel-separated glycoproteins from human cerebrospinal fluid by tandem high-resolution mass spectrometry, *J. Proteome Res.*, **2003**, 2(6): 581-588.

[21] Hakansson, K., Cooper, H. J., Hudgins, R. R., Nilsson, C. L., High resolution tandem mass spectrometry for structural biochemistry, *Curr. Org. Chem.*, **2003**, 7(15): 1503-1525.

[22] Hakansson, K., Chalmers, M. J., Quinn, J. P., McFarland, M. A., Hendrickson, C. L., Marshall, A. G., Combined electron capture and infrared multiphoton dissociation for multistage MS/MS in a Fourier transform ion cyclotron resonance mass spectrometer, *Anal. Chem.*, **2003**, 75(13): 3256-3262.

[23] Kruppa, G., Schnier, P. D., Tabei, K., Van Orden, S., Siegel, M. M., Multiple ion isolation applications in FT-ICR MS: Exact-mass MSn internal calibration and purification/interrogation of protein-drug complexes, *Anal. Chem.*, **2002**, 74(15): 3877-3886.

[24] Palmblad, M., Hakansson, K., Hakansson, P., Feng, X. D., Cooper, H. J., Giannakopulos, A. E., Green, P. S., Derrick, P. J., A 9.4 T Fourier transform ion cyclotron resonance mass spectrometer: description and performance, *Eur. J. Mass Spectrom.*, **2000**, 6(3): 267-275.

[25] Solouki, T., PasaTolic, L., Jackson, G. S., Guan, S. G., Marshall, A. G., High-resolution multistage MS, MS(2), and MS(3) matrix-assisted laser desorption/ionization FT-ICR mass spectra of peptides from a single laser shot, *Anal. Chem.*, **1996**, 68(21): 3718-3725.

[26] Ramsey, R. S., Ramsey, J. M., Generating electrospray from microchip devices using electroosmotic pumping, *Anal. Chem.*, **1997**, 69(6): 1174-1178.

[27] Xue, Q., Foret, F., Dunayevskiy, Y. M., Zavracky, P. M., McGruer, N. E., Karger, B. L., Multichannel microchip electrospray mass spectrometry, *Anal. Chem.*, **1997**, 69(3): 426-430.

[28] Xue, Q. F., Dunayevskiy, Y. M., Foret, F., Karger, B. L., Integrated multichannel microchip electrospray ionization mass spectrometry: Analysis of peptides from on-chip tryptic digestion of melittin, *Rapid Commun. Mass Spectrom.*, **1997**, 11(12): 1253-1256.

[29] Rohner, T. C., Rossier, J. S., Girault, H. H., On-line electrochemical tagging of cysteines in proteins during nanospray, *Electrochem. Commun.*, **2002**, 4(9): 695-700.

[30] Roussel, C., Dayon, L., Lion, N., Rohner, T. C., Josserand, J., Rossier, J. S., Jensen, H., Girault, H. H., Generation of mass-tags by the inherent electrochemistry of electrospray for protein mass spectrometry, *J. Amer. Soc. Mass Spectrom.*, **2004**, in press.

[31] Roussel, C., Dayon, L., Jensen, H., Girault, H. H., On-line cysteine modification for protein analysis: new probes for electrochemical tagging nanospray mass spectrometry, *J. Electroanal. Chem.*, **2004**, 570(2): 187-199.

[32] Dayon, L., Roussel, C., Girault, H. H., On-line electrochemical tagging of free cysteines in peptides during nanospray ionisation mass spectrometry: an overview, *Chimia*, **2004**, 58(4): 204-207.

[33] Dayon, L., Roussel, C., Prudent, M., Lion, N., Girault, H. H., On-line Counting of Cysteine Residues in Peptides during Electrospray Ionisation by Electrogenerated Tags and its Application to Protein Identification, *Electrophoresis*, **2004**, in press.

[34] Figeys, D., Ning, Y., Aebersold, R., A microfabricated device for rapid protein identification by microelectrospray ion trap mass spectrometry, *Anal. Chem.*, **1997**, 69(16): 3153-3160.

[35] Li, J., Kelly, J. F., Chernushevich, I., Harrison, D. J., Thibault, P., Separation and identification of peptides from gel-isolated membrane proteins using a microfabricated device for combined capillary electrophoresis/nanoelectrospray mass spectrometry, *Anal. Chem.*, **2000**, 72(3): 599-609..

[36] Li, J. J., Wang, C., Kelly, J. F., Harrison, D. J., Thibault, P., Rapid and sensitive separation of trace level protein digests using microfabricated devices coupled to a quadrupole-time-of-flight mass spectrometer, *Electrophoresis*, **2000**, 21(1): 198-210.

[37] Li, J. J., LeRiche, T., Tremblay, T. L., Wang, C., Bonneil, E., Harrison, D. J., Thibault, P., Application of microfluidic devices to proteomics research - Identification of trace-level protein digests and affinity capture of target peptides, *Mol. Cell. Proteomics*, **2002**, 1(2): 157-168.

[38] Pfeifer, R. J., Hendricks, C. D., Charge-to-mass relationships for electrohydrodynamically sprayed liquid droplets, *Phys. Fluids*, **1967**, 10(10): 2149-2154.

[39] Delamora, J. F., The Effect of Charge Emission from Electrified Liquid Cones, *J. Fluid Mech.*, **1992**, 243(1): 561-574.

..

Chapter 4. Glycoconjugate analysis

1. Introduction

This chapter presents the collaborative work accomplished with the laboratory of Pr. Jasna Peter-Katalinic, Biomedical Analysis Group, Institute for Physics and Biophysics, University of Münster, Germany. The primary goal of this work was to test the suitability of the microfabricated microsprayers developed in our laboratory for the mass spectrometric analysis of glycoconjugates. As explained in section 4 (Oligosaccharides mass spectrometry), mass spectrometric analysis of carbohydrates and glycoconjugates is challenging in terms of ionisability, spray stability to allow investigation of structural features, and sensitivity. These are the parameters that are studied in sections 5 (Q-TOF analysis of O-glycopeptides and gangliosides) and 6 (FT-ICR analysis of O-glycopeptides). Whereas the goal of this study was purely to assess the potential of microfabricated microsprayers, section 2 describes the basics of oligosaccharides and glycoconjugates biochemistry, and section 3 gives a few elements about the biomedical context of such studies.

2. Basics of glycoconjugate biochemistry

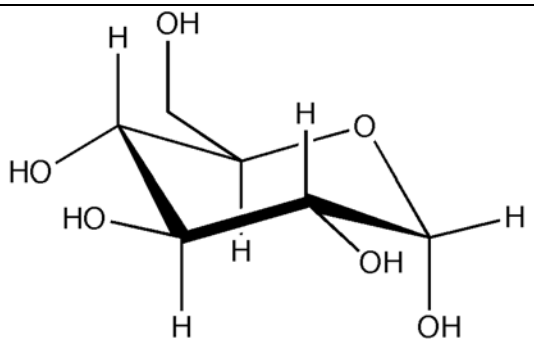
The term glycoconjugate refers to hybrid molecules where carbohydrate chains are covalently linked to other structural entities, such as peptides or proteins, lipids or

extracellular components. The carbohydrate chain itself is often referred to as glycan. Depending of the nature of the molecule the glycan is linked to, glycoconjugates can be classified into:

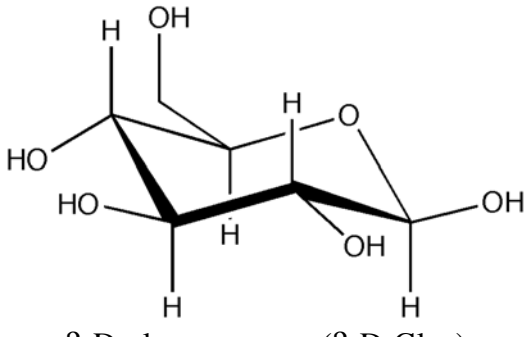
1. glycopeptides and glycoproteins, among which proteoglycans specifically refer to extracellular proteins.
2. glycolipids in which glycans are linked to a lipid moiety.

2.1. Monosaccharides and the glycosidic bond

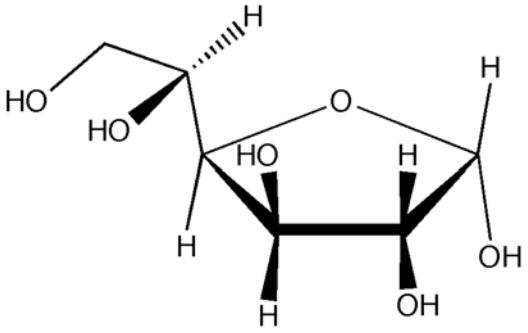
The glycan chain itself is made of monosaccharide units; a monosaccharide is a polyhydroxyaldehyde or polyhydroxyketone of general formula $(CH_2O)_n$ that cannot be hydrolysed into simpler units. Compounds with the same formula can be found in various cyclic forms, namely the pyranose form (6 carbon ring) and furanose form (5 carbon ring):



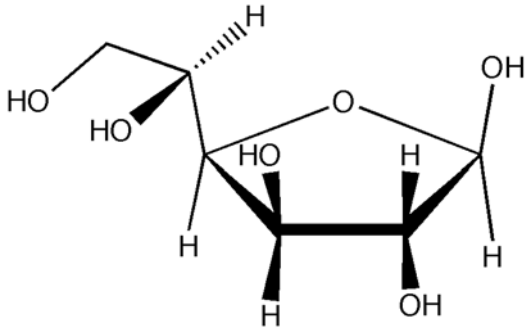
α-D-glucopyranose (α-D-Glc*p*)



β-D-glucopyranose (β-D-Glc*p*)



α-D-glucofuranose (α-D-Glc*f*)



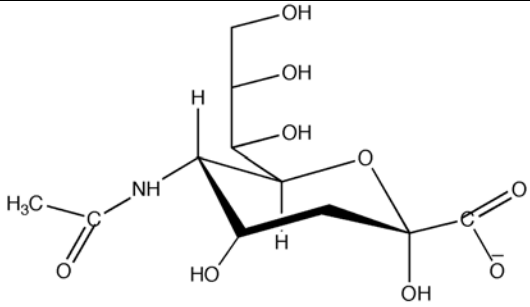
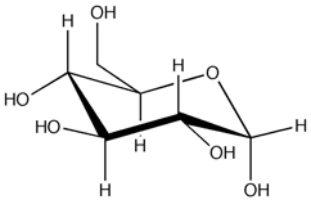
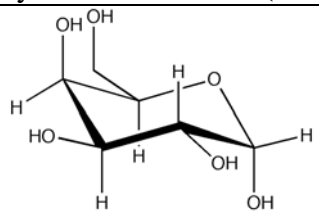
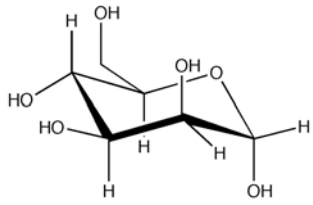
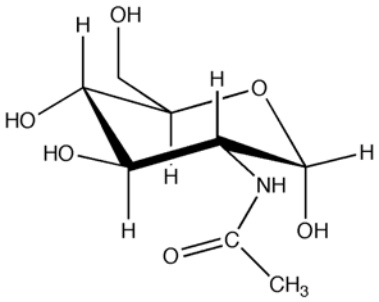
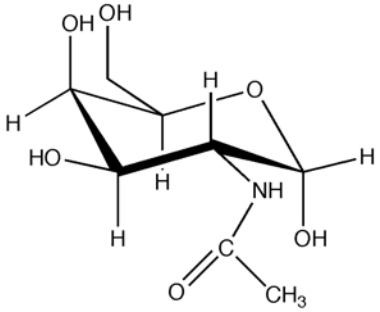
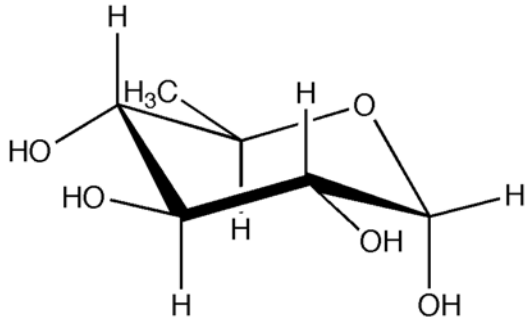
β-D-glucofuranose (β-D-Glc*f*)

Table 5. The four D-isomers of general formula $C_6H_{12}O_6$

Additionally to the number of carbon atoms involved in the ring (pyranose or furanose form), the position of the hydroxyl group of carbon 1 gives additional isomers (α

and β), as well as L and D conformation. For a ring saccharide of general formula $C_6H_{12}O_6$, there are thus 8 possible isomers [1].

The most common monosaccharides in higher organisms are listed in Table 6:

Family	Most commons monosaccharides
Sialic acids (Sia)	 N-acetyl neuraminic acid (Neu5Ac)
Hexoses (Hex)	 glucose (Glc)  galactose (Gal)  mannose (Man)
Hexosamines (HexA)	 N-acetylglucosamine (GlcNAc)  N-acetylgalactosamine (GalNAc)
Deoxyhexoses (dHex)	 fucose (Fuc)

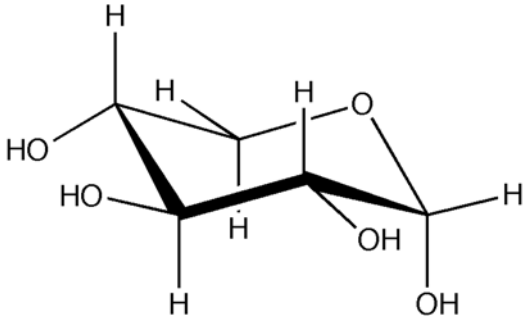
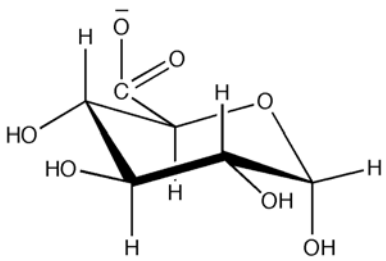
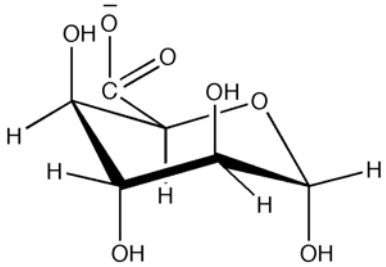
Pentoses (Pen)	 xylose (Xyl)	
Uronic acids	 glucoronic acid (GlcA)	 iduronic acid (IdA)

Table 6. Ten most common saccharide units found in higher organisms (in their α-pyranose form).

Table 6 lists only the most common monosaccharide units in higher organisms. For example, the sialic acid family comprises tens of different molecules; additionally, the set of monosaccharides that can be found in various organisms can be very different [2]. In order to from oligo- or polysaccharides, monosaccharide units assemble through glysosidic linkage, in a manner very similar to the assembly of amino acids into peptides through the peptidic linkage and assembly of bases into DNA through the phosphodiester linkage:

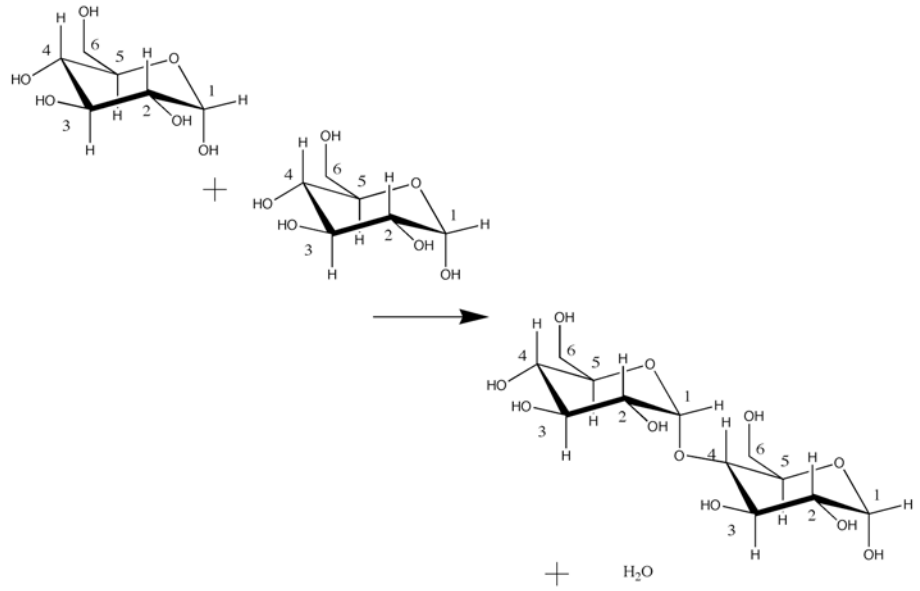


Figure 44. Formation of maltose (D-Glcp-α-(1-4)-D-Glcp-α) from two glucose units.

Figure 44 illustrates the formation of a glycosidic bond between two glucose units to form a maltose molecule. However, the glycosidic bond can involve hydroxyl groups linked to different carbon atoms, including C6. There are thus 15 different disaccharides of primary sequence D-Glc*p*-α-D-Glc*p*-α. Moreover, as all monosaccharides carry several hydroxyl groups, polysaccharides are not necessary linear, but can be branched. Figure 45 shows the classical representation adopted to depict polysaccharides structures:

Sia*p*α2-3Gal*p*β1-4GlcNAc*p*β1-2Man*p*α1 Fuc*p*α1
6 6
Man*p*β1-4GlcNAc*p*β1-4GlcNAc*p*~
3
Sia*p*α2-6Gal*p*β1-4GlcNAc*p*β1-2Man*p*α1

Figure 45. Traditional description of a branched polysaccharide.

Monosaccharide units are from Table 6; *p* specifies that monosaccharide units are in their pyrannose form; the numbers are those of the carbon atom which bear the hydroxyl group involved in the glycosidic bond; greek letters α and β refer to the orientation of the hydroxyl group involved in the formation of the glycosidic bond. By convention, when

unspecified all monosaccharide units are assumed to be in their pyrannose form, and all glycosidic bonds to originate from the C1 hydroxyl (except for sialic acids, which are linked to the C2 hydroxyl). These conventions lead to the simplified traditional representation of polysaccharides:

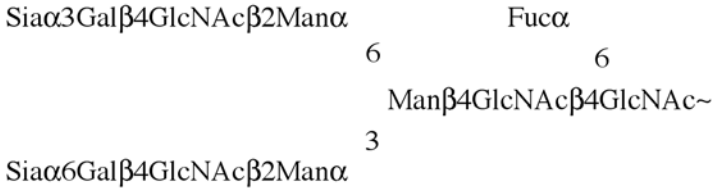


Figure 46. Simplified traditional description of a branched polysaccharide.

Additionnaly, oligo- or polysaccharides can bear chemical modifications (comparable to post-translational modifications in proteins), the most common being phosphorylation, sulfation, methylation, O-acetylation and fatty acylation of hydroxyl groups, N-acetylation and N-sulfation of amino groups.

2.2. Complexity of the sugar code

As shown below, monosaccharides exist in a number of isomeric forms, based on pyranose or furanose forms, configuration L or D and orientation of hydroxyl group α and β. Based on this, the number of different linear n-oligosaccharides would be [3]:

$$E^n \times 2_r^n \times 2_a^n \times \frac{L(L+1)}{n}$$

where E^n represents the number of permutations form the primary sequence; based on 20 common monosaccharides, E^2=400. The term 2^2_r accounts for ring size (pyranose or furanose) and would be 4. The term 2^2_a accounts for anomeric form and would be 4. The last term accounts for possible linkage positions (L being the number of available hydroxyl groups, e.g. 5 for hexoses, 4 for pentoses). For a dissachharide it would be of the order of 12 (15 for hexoses and 10 for pentoses). There are thus a total number of approximately 76800

distincts disaccharides. Even with only D-monosaccharides, as found in the living world, this number would be 19200, compared to 16 dinucleotides and 400 dipeptides.

Similarly there are 768000 distinct linear trisaccharides (not taking into account anomerism); if one considers the case of a trisaccharide chain attached to an aglycon molecule, additional branched isomers must be considered, producing a number of additional isomers (L branching positions for the aglycon molecule, and then L-1 and L-2 respectively for the two branching monosaccharides):

$$E^n \times 2_r^n \times 2_a^n \times L \times \frac{L(L-1)}{2} \times \frac{(L-1)(L-2)}{2}$$

For a branched trisaccharide, it would add an additional 9922500 possibilities (not taking anomerism into account), for a total (linear plus branched) close to 10^7 different molecules, compared to 64 trinucleotides and 8000 tripeptides.

2.3. Glycopeptides and glycoproteins

Polysaccharide (glycan) chains are linked to proteins via asparagines (Asn/N) on consensus sequences NX-S/T; these are N-glycans. N-glycans are divided into three subclasses:

1. *complex type* containing GalNAc, GlcNAc, Man, Gal and Neu5Ac residues
2. *high mannose type* containing GlcNAc and a high number of Man
3. *hybrid type* containing both complex and high mannose types strctures.

N-glycosylation has been widely studied and biosynthesis pathway are reasonnably well understood (see below).

Glycan chains can also be linked to proteins through linkage to serine or threonine residues; these are O-glycans. No consensus sequence has been found for O-glycosylation. The modification of serine or threonine residues occurs by addition of a GalNAc residue. Eight core structures have been identified, that are presented in Table 7:

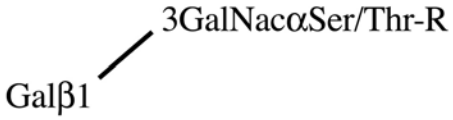
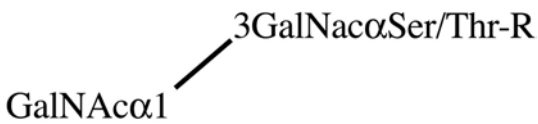
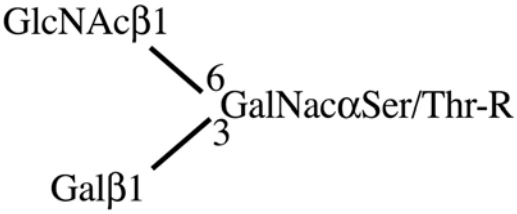
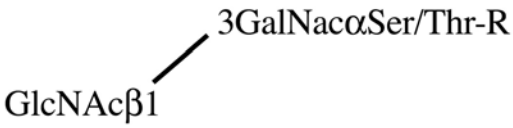
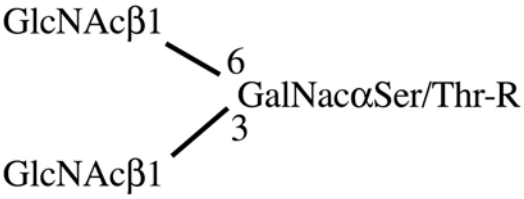
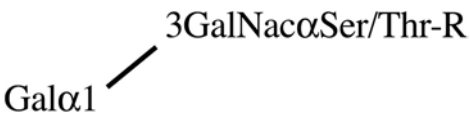
 core 1	 core 5
 core 2	 core 6
 core 3	 core 7
 core 4	 core 8

Table 7. The eight core structures in O-glycosylation.

2.4. Glycolipids

Glycolipids are glycoconjugates that comprise one or more saccharide unit glycosidically linked to a hydrophobic moiety such as acylglycerol, ceramide or prenyl phosphate. According to the type of lipid moiety linked to the glycan chain, glycolipids are classified as:

a. glycoglycerolipids: glycolipids containing one or more glycerol residue :

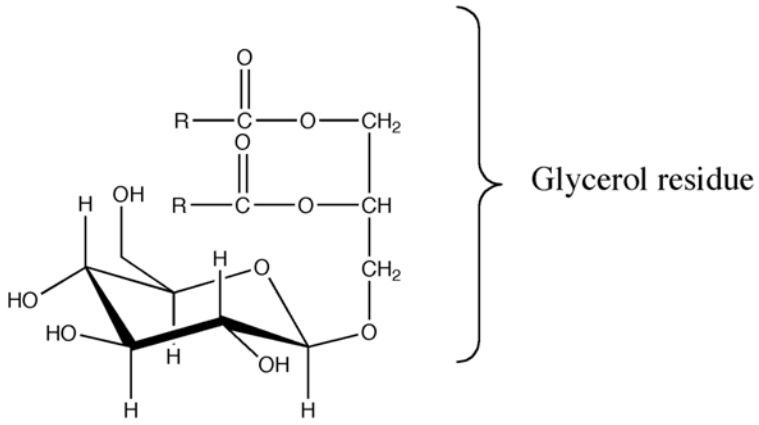


Figure 47. Typical structure of a glycerolipid.

b. glycophosphatidylinositols (GPI): oligosaccharides glycosidically linked to the inositol moiety (including those bearing different modifications such as *O*-acyl, *O*-alkyl… on their glycerol or inositol residues). While the diversity of GPIs is only being unravelled, many appear to share a common "core" (see Figure 48).

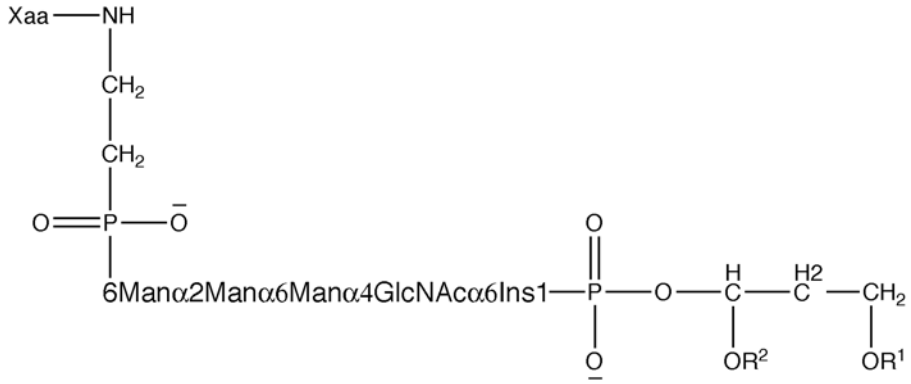


Figure 48. Typical glycophosphatidylinositol (GPI).

GPIs covalently attached to peptides or proteins (usually through the C-terminus via an amide linkage) are called GPI anchors.

c. glycosphingolipids (GSL) are carbohydrate-containing derivatives of a sphingoid or ceramide. The most prominent lipid anchor in higher organisms is a ceramide,

which consists of sphingosine substituted at its NH_2 group by fatty acids of various lengths. There are two types of GSL: neutral, comprising neutral sugar moieties, and acidic comprising groups such as sialic acids, uronic acids… Syalilated GSL called gangliosides carry sialic acid linked to C3 or C6 of Gal or GalNAc of the core oligosaccharide.

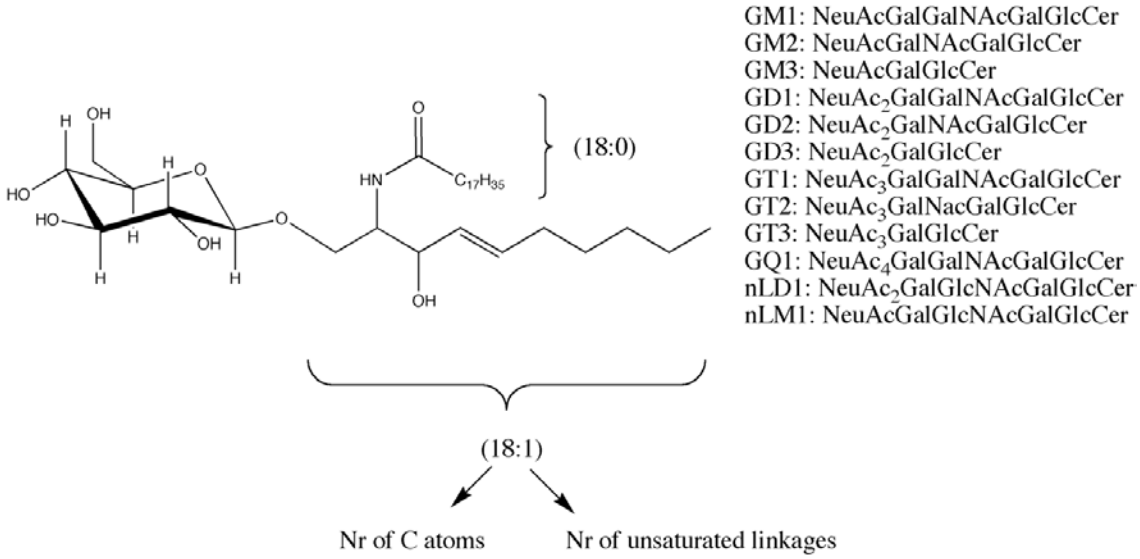


Figure 49. Simple glycosphingolipid Gal-Cer. Monosaccharide composition of major ganglioside species are shown on the right.

The ceramide moiety potentially bears a certain heterogeneity, that is clarified by the indication (d18:1/18:0), which indicates the number of carbon atoms in each branch of the ceramide as well as the number of unsaturated bonds. Nomenclature of glycolipids has been recently published by the Union for Pure and Applied Chemistry [4].

3. Physiology and biomedical aspects

Glycoconjugates are ubiquitous through the living world, but tend to be more complex and more abundant in higher organisms. Glycans definitively play a role in cell structure, especially when attached to matrix molecules such as collagen and proteoglycans.

They may also have a general protective role against proteolysis, as is well known for immunoglobulins. Protein glycosylation can also induce an on/off effect, such as for the β-human chorionic gonadotrophin (βHCG), which is unable to activate adenylate cyclase when deglycosylated [5], or fine tuning of an enzymatic function or receptor affinity. Glycans can also act as low affinity binders in the extracellular domain: glycosaminoglycans bind chemokines on the cell surface and facilitate the formation of chemokine gradients that direct cell migration in inflammation [6].

Protein glycosylation also plays a recognition role, such as in protein clearance from plasma [7], in cell-cell and cell-matrix interactions: for example the glycosylation of selectin ligands play a crucial role in the regulation of critical interactions among blood cells [8, 9]. As key elements that take part in cell-cell communication, glycans are also recognised as key diagnostic and therapeutic targets in cancer [10]. These ubiquitous and versatile roles of glycans in most physiological processes imply that they indeed take part in many physiological disorders and diseases.

3.1. Congenital disorders of glycosylation (CDGs)

One class of diseases that is directly linked to glycosylation alterations are congenital disorders of glycosylation (CDGs), formerly named carbohydrate-deficient glycoprotein syndromes. CDGs are inherited disorders affecting the assembly or processing of glycans on glycoconjugates. One difficulty in the diagnosis of CDGS is that their clinical spectrum is very variable, ranging from strabismus or ataxia to lethal multisystemic disorders such as cardiomyopathy, nephritic syndromes, hepatic fibrosis…, statomotor and mental retardation being very frequent. They can be considered as orphan diseases as only a few hundreds patients are affected worldwide, and little research is done: the first molecular description of a disorder in the biosynthesis of *N*-glycans has been reported in 1995 only

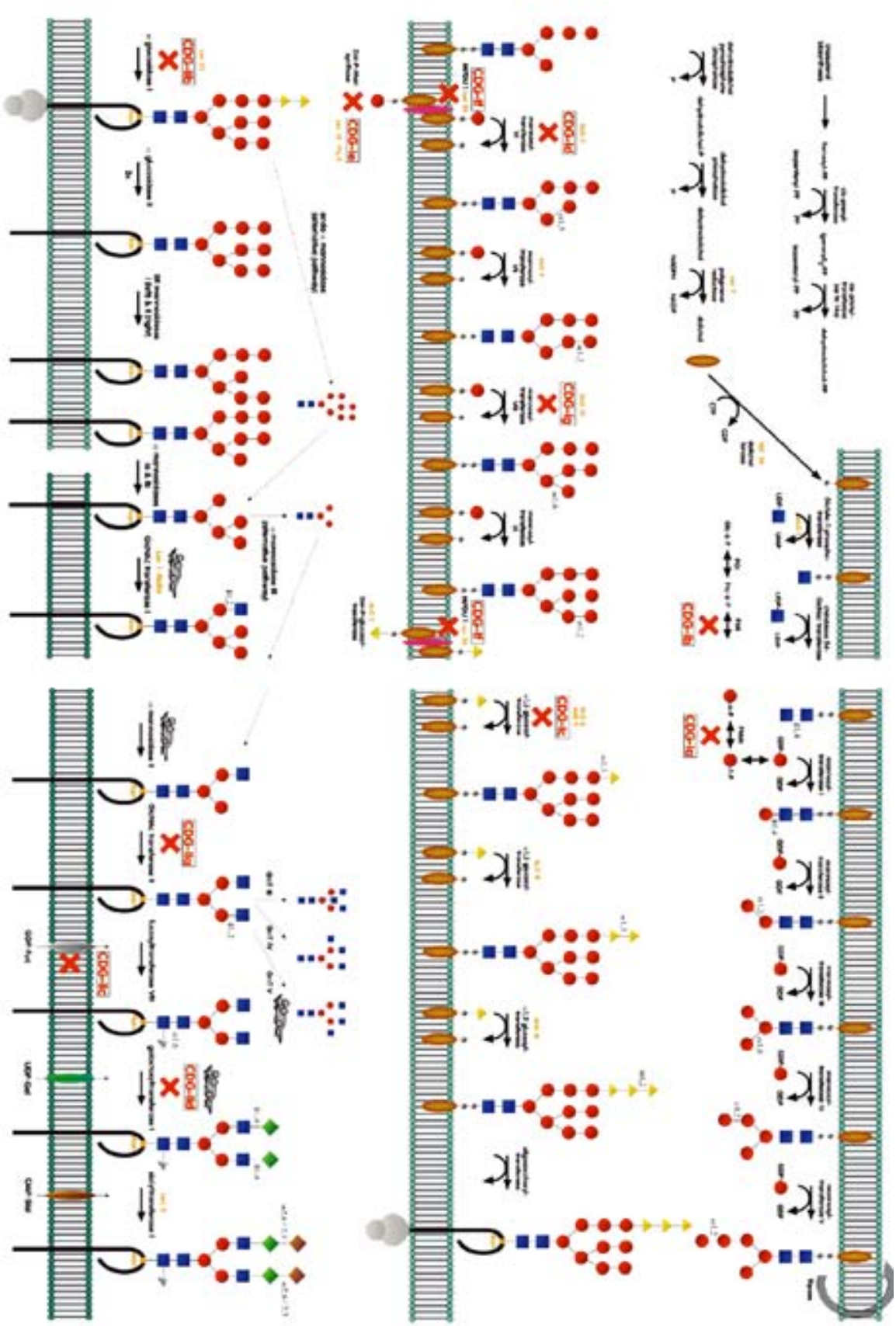
[11]. Molecular diagnosis is mainly based on the isoelectric focusing pattern of serum transferrin: however, this diagnostic procedure is only indirect, and not all CDGs result in alterations of transferrin [12].

Most types of CDGs discovered so far relate to defects in the *N*-glycosylation pathway of glycoproteins. But it should be pointed out that *N*-glycosylation is relatively well-known, whereas little is known about the *O*-glycosylation biosynthetic pathway; moreover, 20% of diagnosed CDGs have not received molecular explanation yet. It is thus highly probable that new types of CDGs will be identified in the years to come.

3.2. *N*-linked glycoproteins

Protein translation occurs on ribosomes either in the cytoplasm or on the endoplasmic reticulum (ER). Upon the start of translation, the N-terminal part of the synthesised protein is translocated within the ER. The building of an *N*-linked glycan structure onto asparagines residues start with the synthesis of a dolichol-saccharide precursor in the ER, which is brought in the vicinity of the protein and the saccharide is enzymatically transferred from its dolichol anchor to an asparagine residue. A great number of biosynthesis steps then build the glycan chain on this first saccharide. The protein is then transferred to the Golgi by vesicle transport where further carbohydrate heterogeneity is created by a whole enzymatic apparatus. An overview of this biosynthetic pathway is shown in Figure 50.

Figure 50. (next page) Scheme of the biosynthetic pathway of N-glycosylation. Reprinted with permission from [12].



This pathway involves tens of enzymes (most of then are protein complexes); any enzymatic alteration (lack of activity, hyperactivity, wrong activity) can lead to the production of misglycosylated proteins that will not be able to assume their physiological functions. Documented CDGs are indicated by red crosses along the biosynthetic pathway in Figure 50.

3.3. *O*-linked glycoproteins

Whereas *N*-glycosylation occurs cotranslationally in the ER and further in the Golgi, *O*-glycosylation occurs posttranslationally and occurs downstream of the protein production. In the most common type of *O*-glycosylation (mucin-type-*O*-glycosylation), an N-acetylgalactosamine (GalNAc) is attached to Ser/Thr of the protein [13]. Alternatively, *O*-GlcNAc addition on Ser/Thr , direct glucosylation, fucosylation and galactosylation, and *O*-mannosylation also exist [12, 14, 15].

Many glycan processing enzymes are specific to *O*-glycosylation, but there are also some documented overlapping between *N*- and *O*-glycosylation biosynthetic pathways.

3.4. Schindler disease

Schindler or Kanzaki disease is defined as an α-*N*-acetylgalactosaminidase (α-NAGA) deficiency [16], a genetic autosomal recessive trait inherited metabolic disorder, characterised by deficient activity of the lysosomal enzyme α-NAGA and consequent abnormal urinary excretion of glycoconjugates [17]. A deficiency of the activity of α-NAGA should theoretically result in the accumulation of undegraded glycoconjugates terminating with the GalNAc residue. However, major urinary excretion products have been found to be sialylated glycosylated amino acids including NeuAcα2-3NeuAcβ1-3[NeuAcα1-2]GalNAcα1-*O*-Ser/Thr and NeuAcα2-3Galβ1-3GalNAcα1-*O*-Ser/Thr [18].

Therefore the assessment of urinary glycopeptides present in both healthy individuals and patients suffering from Schindler disease is of major importance for a better understanding of the disease mechanism.

3.5. Glycans and cancer

Glycobiologists have known for a long time that the structures of glycans that decorate eukaryotic cell surfaces, change with the onset of cancer [19]. Changes in glycosylation include under-, over-, and neoexpression of glycans. Most often these modifications are due to changes in expression levels of glycosyltransferases in the Golgi [10], which both affects *N*- and *O*-glycoproteins. Another common feature is the over- or altered expression of glycolipids, in particular of gangliosides. Gangliosides are sialic acid containing glycosphingolipids (see paragraph 2.4 above), present in the cell membrane of all vertebrates and especially in the cells of the nervous system. Changes in their composition and content have been observed during physiological growth and differentiation as well as in neoplastic cell transformation. A decrease in the regular ganglioside profile and an increase in the structures detected only in small amounts in normal brain tissue have been shown in primary brain tumours [20, 21]. This features demonstrates the direct correlation between ganglioside composition and histological type and grade of the tumours and that changes in ganglioside expression may serve as biochemical markers in early histopathological diagnosis, grading and prognosis of tumours. Additionally, for tumour treatment, one strategy may be to use ligands that specifically bind the invading cells for which gangliosides are a potential group of tumour-specific targets.

4. Oligosaccharides mass spectrometry

This section is mainly based on the insightful review by Zaia et al [22]. It aims to explain the general principles of oligosaccharide mass spectrometry as well as its specificities compared to peptide and protein mass spectrometry.

4.1. Ionisation of carbohydrates

Carbohydrates have been studied by both electrospray and matrix-assisted laser desorption/ionisation mass spectrometry. Electrospray ionisation usually results in much less intense signals than that obtained form peptide solutions of the same concentration. The main reason for this low sensitivity is the hydrophilic nature of oligosaccharides. It is now well-established that hydrophobicity of analytes play a crucial role in determining their spectrum intensity [23, 24]. There can easily be one or two orders of magnitude differences in the mass spectrum due to differences in hydrophobicity. This has led to derivatisation strategies to decrease oligosaccharides hydrophilicity, e.g. through permethylation or peracetylation. Derivatisation strategies will not be described herein because experiments described below concerns only non-derivatised oligosaccharides. For more details, the reader is referred to references [22, 25-27].

Besides derivatisation, nanoelectrospray (nESI) has received particular attention for the analysis of oligosaccharides. Nanoelectrospray produces much smaller charged droplets that electrospray at high flow rates [28]: in ESI, typical droplet size is about 1 μm, whereas in nESI, primary droplets have diameters of the order of 100 nm. This thus lead to an increase of surface-to-volume ratios typically ten times higher. The influence of surface affinity in analyte charging is thus believed to be less critical [29]. Figure 51 shows ESI mass spectra of maltopentaose and bovine insulin at equimolar concentration in microESI and nanospray. In the microESI spectrum (Figure 51a), maltopentaose appears as a sodiated

peak represented only 5% of the insulin base beak, whereas in nanospray, maltopentaose and insulin give peaks of comparable intensity.

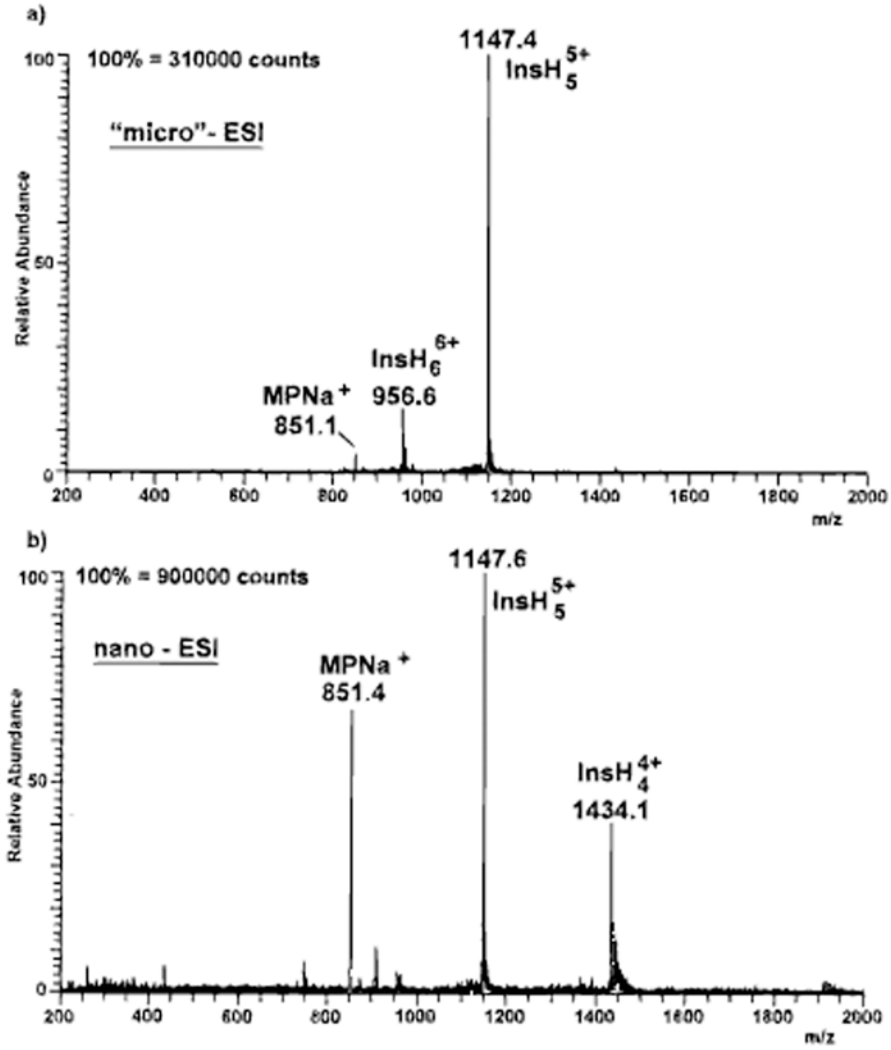


Figure 51. ESI mixture analysis of maltopentaose and insulin (both 5×10^{-6} M) with conventional forced-flow ("micro") (a) and nanospray (b). The averaged absolute intensities (counts) for the base peaks are indicated. Reprinted with permission from [30].

On the other hand, oligosaccharides have also been analysed by matrix assisted laser desorption/ionisation mass spectrometry (MALDI-MS). Whereas MALDI-MS appears to be well-suited for heterogeneous mixtures (such as glycans released from glycoproteins) due to much less critical competition for ionisation between analytes (known as "ion suppression" in ESI), it suffers from a major drawback: more energy is transferred to analyte molecules

compared to ESI, which results in fragmentation of labile moieties such as sialic acids [29]. One general way to limit energy deposition during ionisation processes is to use atmospheric pressure ionisation sources: droplets and desolvated ions are then collisionally cooled by the ambient gas. But atmospheric MALDI (AP-MALDI) sources have been introduced only recently [31-33] and only limited work has been done for AP-MALDI analysis of oligosaccharides to date [34].

4.2. Nomenclature for fragmentation of glycoconjugates

The nomenclature for fragmentation of glycoconjugates is relatively similar to that for peptides and has been established by Domon and Costello in 1988 [35]. Fragment ions that contain a non-reducing terminus are labelled with upper case letter from the beginning of the alphabet (A, B, C) and those that contain the reducing end of the oligosaccharide or the aglycon are labelled with letters from the end of the alphabet (X, Y, Z). Subscripts indicate the cleaved ions. Additionally, A and X ions are produced by cleavage across the glycosidic ring, and are labelled by assigning each ring bond a number from the ring oxygen atom and counting clockwise, as shown in Figure 52.

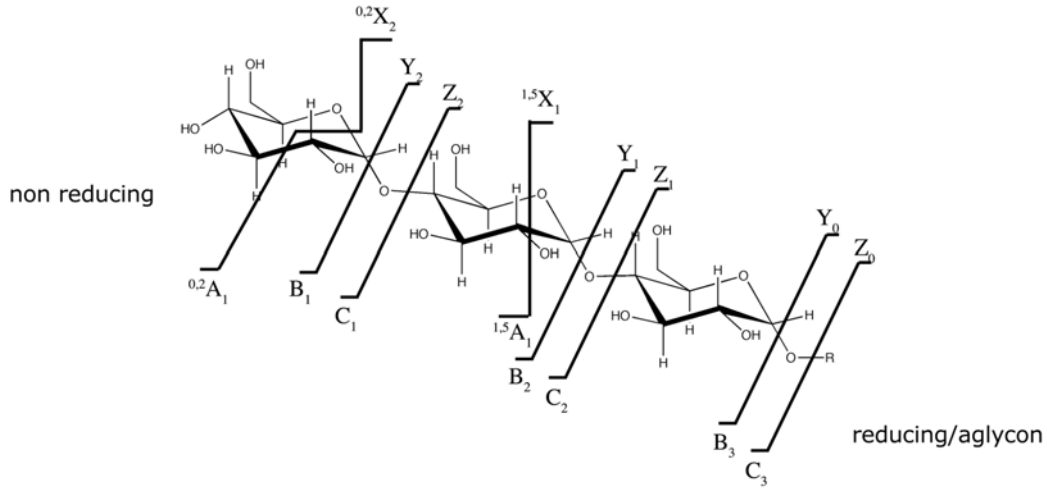


Figure 52. Nomenclature for glycoconjugate product ions generated by tandem MS. (please note that for the ease of representation, the angles of the glycosidic bond were not respected).

4.3. Challenges of oligosaccharide mass spectrometry

Whereas mass spectrometry is the most powerful technique applied to date for the elucidation of oligosaccharide structures, it still does not allow easy determination of oligosaccharide sequence; this is explained by the isomeric nature of monosaccharide units, as shown in Table 8:

Compound	Monoisotopic mass	Average mass
Neu5Ac	308.0987	308.26247
Glc	180.06339	180.15588
Gal	180.06339	180.15588
Man	180.06339	180.15588
GlcNAc	221.08994	221.2078
GalNAc	221.08994	221.2078
Fuc	164.06847	164.15648
Xyl	150.05282	150.1299
GlcA	193.03538	193.13201
IdA	193.03538	193.13201

Table 8. Monoisotopic and average masses of monosaccharides listed in Table 6.

Mass measurement alone does not allow discrimination between Glc, Gal and Man, or between GlcNAc and GalNAc, for example. Collision-induced-dissociation (CID)

usually results in preferential cleavage of glycosidic bonds, leading to profusion of B and Y ions. To a less extent, ring-cleavage product ions can also be produced, but tend to disappear for larger ions (due to conformational cooling that can occur preferentially to ring cleavage, which implies the cleavage of two covalent bonds); this effect can be seen as similar to the difference in CID spectra of proteins compared to peptides. But at the same time, product ions coming from glycosidic bond cleavage can only give sequence information, but no structural information, as it does not allow the discrimination between isomers (such as the different hexoses or hexosamines). Only ring-cleavage product ions provide true structural information, such as isomer discrimination or branching information).

5. Q-TOF analysis of O-glycopeptides and gangliosides

5.1. Material and methods

Chemicals

Analytical grade methanol was obtained from Merck (Darmstadt, Germany) and used without further purification. Distilled and deionised water (Mili-Q water systems Millipore, Bedford, MA, USA) was used for the preparation of the sample solutions. Sample solutions were dried in a SpeedVac SPD 111V system from Savant (Düsseldorf, Germany) and centrifuged in an Eppendorf 5415 C centrifuge (Hamburg, Germany).

Biological samples

Mixture of O-glycosylated amino-acids and peptides from normal human urine.

The complex mixture of O-glycosylated amino-acids and peptides investigated in this work was obtained and partially purified previously from urine of a healthy subject.

The extraction, separation and purification procedure was described by us previously [17, 36]. The sample is denoted Ty. Stock solutions of the Ty mixture at 1 mg/ml were prepared by dissolving the dried material in 100% MeOH and freezing it at -200C. Dilution of the stock solutions in pure MeOH yielded the working aliquots at the concentration of 5.0 and 1.25 pmol/ml calculated for an average molecular weight of 2000 Da.

GT1 ganglioside fraction from normal adult cerebrum

Normal adult (45 years of age) human cerebrum, without pathological signs according to morphoanatomical and histopathological examination was obtained from the Department of Forensic Medicine, Faculty of Medicine, University of Zagreb, Croatia.

The brain tissue was weighed and stored at -20 °C until the extraction procedure. The native fraction of GT1 gangliosides was extracted, isolated and purified as described in detail elsewhere [37]. The ganglioside sample, denoted further GT1, was dried in Speed Vac SPD 111V system (Savant, Düsseldorf, Germany). For microchip-ESI MS analysis, the stock solution of the sample at approximately 1 mg/ml was prepared by dissolving the dried material in methanol to be stored at–200°C. Dilution of the stock solution in 100% methanol yielded the working aliquots at the concentration of approximately 5 pmol/ml of ganglioside extract.

Mass spectrometry

Mass spectrometry was performed on an orthogonal hybrid quadrupole time-of-flight mass spectrometer (QTOF™ Micromass, Manchester, U.K.) in Micromass Z-spray geometry. QTOF mass spectrometer is interfaced to a PC computer running the MassLynx software under Windows N.T. system to control the instrument, acquire and process MS data. Nitrogen was used as desolvation gas and the source block temperature was kept at 800°C. The sampling cone potential and voltage applied on the microsprayer were adjusted

to provide an optimal ionization, minimal in-source fragmentation and a high and sustained electrospray signal.

Tandem mass spectrometry was performed by collision-induced dissociation (CID) at low energies using Ar as a collision gas. In order to obtain a maximum coverage of sequence ions, the collision energy was kept at 40 eV for glycopeptide sequencing and adjusted during the ongoing experiment from 40 to 70 eV for sequencing of GT1 species.

All mass spectra were acquired in the negative ion mode. For mapping and sequencing of glycoconjugates, the negative ion detection was shown to be advantageous in comparison with the positive ion mode [22, 38]. In particular, the assignment of already known or previously unknown components in biological mixtures is easier due to a lower degree of heterogeneity caused by cation attachment to molecular and/or fragment ions [22, 37-40]. After acquisition, the spectra were calibrated using sodium iodide as a calibrant.

Microchip system

The microchip fabrication has been described previously [41, 42]. Basically, a photoresist is patterned on a 75 µm thick, copper-coated polyimide foil through a printed slide acting as a mask. Photoresist is then developed, and chemical etching is then used to remove the deprotected copper where microchannels are to be patterned. Polyimide is plasma-etched to the desired depth. The final microchannels are 120 µm wide, 45 µm deep (nearly "half moon" cross section), with gold-coated microelectrodes placed at the bottom of the microchannel. A 35 µm polyethylene/ polyethylene terephthalate is laminated to close the channels. Finally, one end of each channel is manually cut in a tip shape, so that the outlet of the microchannel is located on the edge of the chip, as described by Gobry et al [43]. For sample dispensing, a reservoir is pasted over the inlet of the microchannel, or the chip is sandwiched in a home-made chip holder with an integrated reservoir.

The whole chip/reservoir assembly was mounted to the QTOF MS. In order to realise the electrical contact to the ESI power supply, the ESI QTOF source was removed and the chip system was directly connected to the ESI high voltage plate, which is a fixed part of QTOF conventional Micromass ESI source. Exchange between the original source and chip system interface did not claim for any definitive dismantling or special mechanical modifications to either of the original assembly and no further modifications on the TOF/MS analyzer were necessary. The position of chip emitter was adjusted in the vicinity of the entrance hole of the sampling cone by the source assembly, which can be manipulated in x, y and z direction via micrometer screws. The microsprayer tip was placed at a distance less than 5 mm, although fine positioning of the microsprayer tip turned out not to be a critical parameter.

The electrical contact was ensured by a conductive wire with one terminal connected to the chip electrode and the other fixed on the ESI high voltage plate. The spray could be initiated at values of 2-3 kV, in the negative ion mode, applied to the nanoESI plate and 80-100 V applied to the sampling cone without the need of nebuliser gas. Figure 53 shows the electrospray generated from the microfabricated microsprayer.

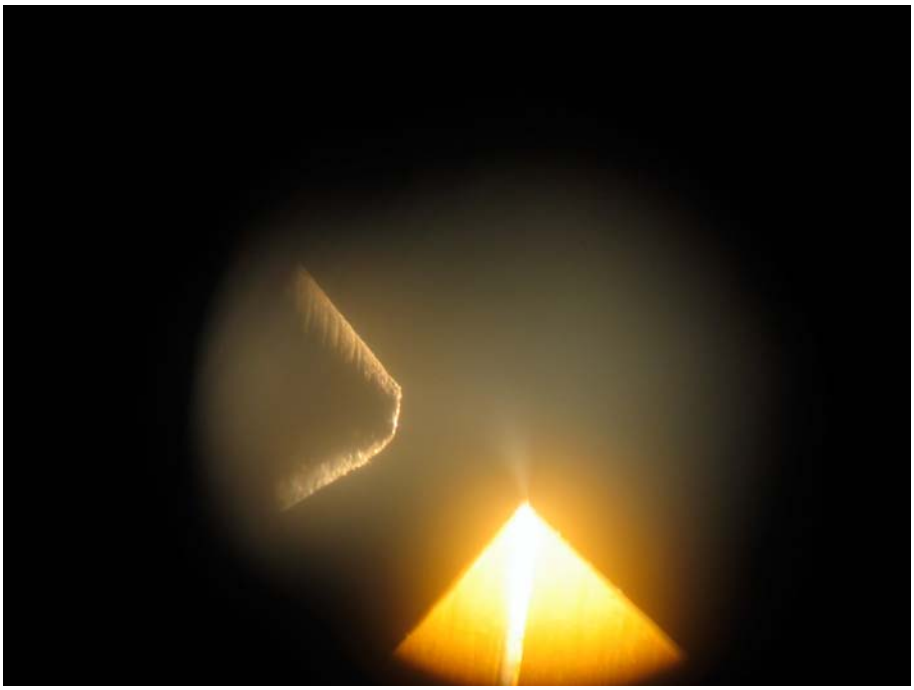


Figure 53. Electrospray generated from the microfabricated microsprayer in the Z-spray source of the QTOF. The sampling cone is visible on the left, whereas the counter electrode is positioned in front of the chip (top of the picture, not visible).

The chip microchannel was flushed with methanol between two consecutive experiments with the same sample. For all experiments, desolvation at 800°C under drying gas was effective, so that the chip unit could be left connected to the mass spectrometer during the channel rinse.

For each sample a fresh chip was used thus any contamination was prevented. The flow rate generated under the employed source and solution parameters was about 200 nl/min [43, 44].

5.2. Results and discussion

Determination of O-GalNAc-Ser/Thr expression in normal human urine

In human urine, carbohydrates are catabolic products excreted as either oligosaccharides or glycopeptides and their concentration, amount and structure is known to

vary under different physiological and/or pathological conditions. For this reason, screening and identification of glycopeptide expression in human urine is of major biological importance.

Ty mixture of *O*-GalNAc glycosylated amino acids and peptides extracted from normal human urine was dissolved in pure methanol to a concentration of 5 pmol/µl. An aliquot of 10 µl was dispensed into a reservoir pasted over the inlet of the chip microchannel. The negative ion mode electrospray process was initiated at 2.8 kV applied to the ESI high voltage plate and 100 V potential of the sampling cone. Under these well-defined conditions a constant and stable spray accompanied by a high intensity of the total ion current was generated. In Figure 54a, the (-)chipESI/QTOF MS1 of the Ty sample at 5 pmol/µl concentration is presented.

The signal was acquired over 20 scans (40 seconds) which at the flow rate of about 200 nl/min is equivalent to a sample consumption of 0.66 pmol. The spectrum combined over 20 scans exhibits a high signal/noise ratio and a number of 26 different saccharide components. Fifteen species detected as singly and/or doubly charged ions are expressing *O*-GalNAc-Ser/Thr core-motif extended by either sialylation or fucosylation. The mixture is dominated by Ser- and Thr- linked disialo saccharides with chain lengths ranging from tetra- to octasaccharide. The monosialylated glycopeptides are also well represented in the spectrum in Figure 54a, being detected as singly charged ions at m/z 760.12, NeuAcHexHexNAc-Ser, and 774.12, NeuAcHexHexNAc-Thr. Mono- and disialylated free oligosaccharides are observed as singly charged ions at m/z 470.03 corresponding to disaccharide NeuAcHex, m/z 673.12 assigned to NeuAcHexHexNAc trisaccharide and m/z 964.16 assigned to NeuAc2HexHexNAc tetrasaccharide. The tetrasaccharide is observed also as a doubly charged ion at m/z 481.09.

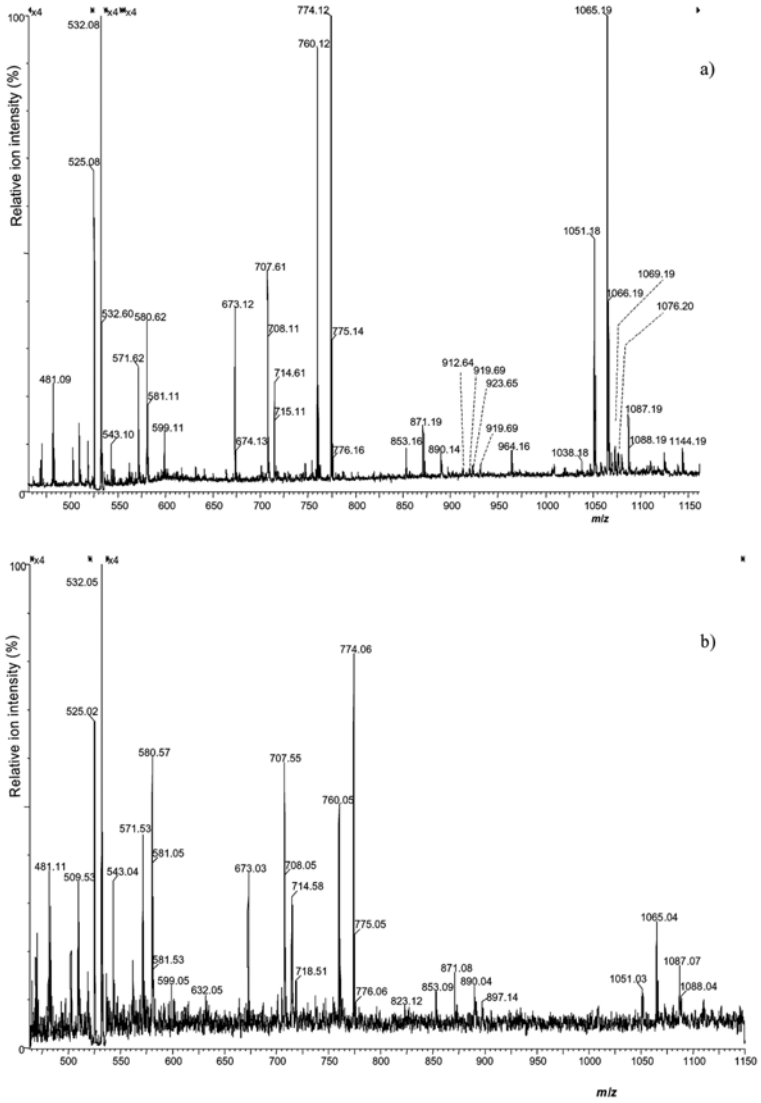


Figure 54. (-) Microchip ESI QTOF MS of the Ty mixture of O-glycosylated amino acids and peptides from human urine. ESI voltage 2.8 kV; Sampling cone potential 100V; Signal acquisition 20 scans. Solvent : MeOH; Average sample concentration: a) 5 pmol/µl; b) 1.25 pmol/µl.

Interestingly, two new structures, not detectable before by any MS-related methods are visible as doubly charged ions at *m/z* 1069.19 and 1076.20. The species could originate from glycopeptides with extended and further modified chains.

In order to test the limit of microchip sensitivity for glycoconjugate detection, Ty solution was diluted in pure methanol yielding the aliquot at 1.25 pmol/µl concentration. In Fig. 1b, the (-) microchip ESI/QTOF MS1 of the Ty sample at 1.25 pmol/µl concentration is

presented. Setting the same source parameters, the signal was acquired over 20 scans, which is equivalent to 0.16 pmol sample consumption for this experiment. Even under these restrictive concentration conditions, a fair signal/noise ratio and 13 different components in the mixture could be identified.

In Table 9 the assignment of the major species present in Ty sample and detected by microchip ESI/QTOF MS1 is given.

m/z (monoisotopic)	Charge state	Putative Structure	5 pmol/µl	1.25 pmol/µl
470.05	1-	NeuAcHex	+	+
481.11	2-	$NeuAc_2HexHexNAc$	+	+
525.08	2-	$NeuAc_2HexHexNAc$-Ser	+	+
532.08	2-	$NeuAc_2HexHexNAc$-Thr	+	+
543.10	2-	$NeuAc_2HexHexNAc$-Thr (Na)	+	+
571.62	2-	$NeuAc_2HexHexNAc$-Thr-Pro -H_2O	+	+
580.62	2-	$NeuAc_2HexHexNAc$-Thr-Pro	+	+
673.12	1-	NeuAcHexHexNAc	+	+
707.61	2-	$NeuAc_2Hex_2HexNAc_2$-Ser	+	+
714.61	2-	$NeuAc_2Hex_2HexNAc_2$-Thr	+	+
760.12	1-	NeuAcHexHexNAc-Ser	+	+
774.12	1-	NeuAcHexHexNAc-Thr	+	+
853.16	1-	NeuAcHexHexNAc-Thr-Pro -H_2O	+	+
871.19	1-	NeuAcHexHexNAc-Thr-Pro	+	+
890.14	2-	$NeuAc_2Hex_3HexNAc_3$-Ser	+	+
897.17	2-	$NeuAc_2Hex_3HexNAc_3$-Thr	+	+
912.64	2-	$NeuAc_2Hex_2HexNAc_4$-Ser -H_2O (Na)	+	
919.69	2-	$NeuAc_2Hex_2HexNAc_4$-Thr -H_2O (Na)	+	
923.65	2-	$NeuAc_2Hex_2HexNAc_4$-Ser -H_2O (2Na)	+	
930.67	2-	$NeuAc_2Hex_2HexNAc_4$-Thr -H_2O (2Na)	+	
964.16	1-	$NeuAc_2HexHexNAc$	+	
1051.18	1-	$NeuAc_2HexHexNAc$-Ser	+	+
1065.19	1-	$NeuAc_2HexHexNAc$-Thr	+	+
1069.19	2-	n.a.	+	
1076.20	2-	n.a.	+	
1087.19	2-	n.a.	+	+
1144.19	1-	$NeuAc_2HexHexNAc$-Thr-Pro -H_2O	+	

Table 9. Compositional mapping of the purified native Ty mixture of glycopeptides from normal human urine as detected by (-) microchip-ESI-QTOF MS. +: ions detected at the given concentration. n.a.: not assigned.

An interesting feature of the spectra in Figure 54 is that at 100 V cone potential, the charge distribution is not shifted towards lower values as observed in the capillary-based ESI QTOF MS experiments, the intensity of the signals corresponding to doubly charged ions being the highest ones. Moreover, the advantage of the microchip ESI MS regarding the minimization of the in-source decay is clearly illustrated in Figure 54 where the *in-source* fragmentation of molecular ions is not observed at such a high value of the sampling cone potential.

To investigate the possibility of performing rapid and accurate glycopeptide sequencing by MS/MS using the polymer chip for sample infusion, the doubly charged ion at *m/z* 532.08, assigned according to the *m/z* value to the already known structure of NeuAc2HexHexNAc-Thr, was isolated and submitted to low-energy CID (-) microchip ESI-QTOF MS/MS. Collision energy, collision gas pressure and precursor ion isolation parameters were carefully adjusted to provide the full set of structural information upon the molecule. The product ion spectrum of the doubly charged ion at *m/z* 532.08 is depicted in Figure 55. The spectrum was obtained after 30 scans (1min) of signal acquisition, resulting in a sample consumption of 1.23 pmols.

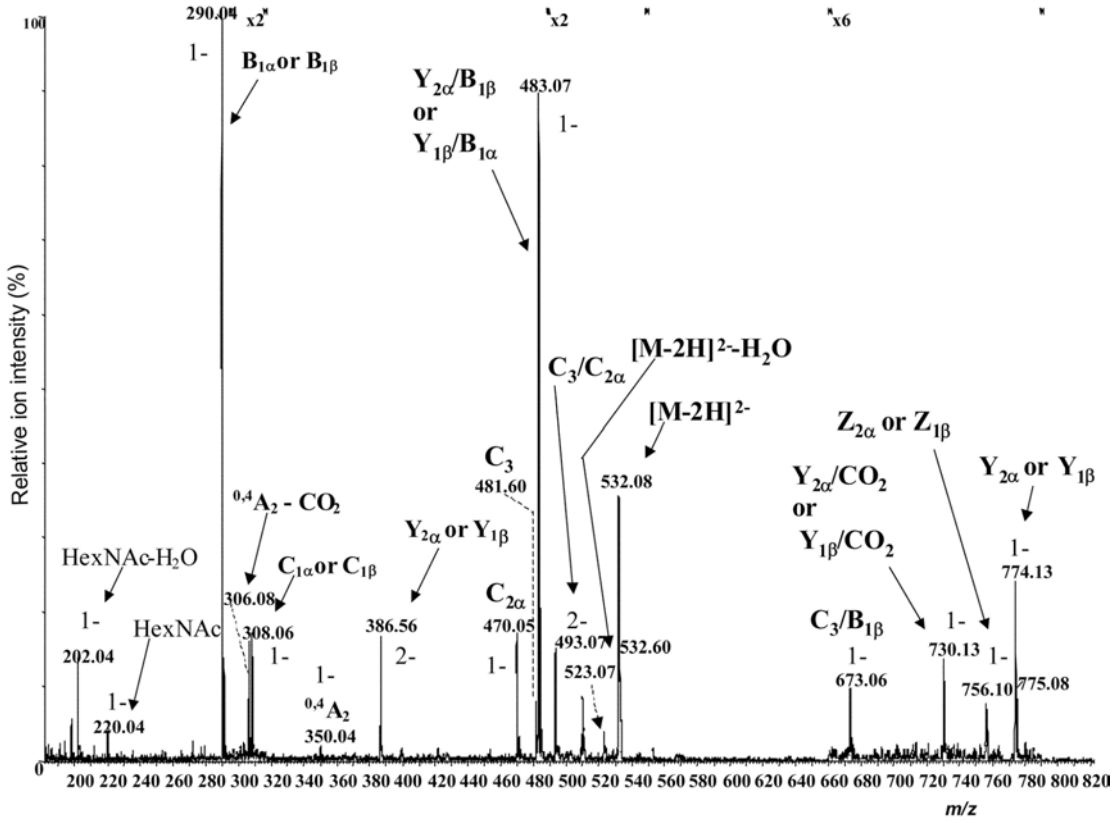


Figure 55. (-) Microchip ESI QTOF MS/MS of the NeuAc2HexHexNAc-Ser doubly charged ion at m/z 532.08. ESI voltage 2.8 kV; Sampling cone potential 100V; Solvent : MeOH; Average sample concentration 5 pmol/μl; Collision energy 40 eV; Signal acquisition 30 scans; Sample consumption 1.23 pmols. The nomenclature for assignment of fragment ions is according to published recommendations by Domon and Costello [35].

The product ion spectrum exhibits a high signal/noise ratio and reveals a significant number of fragment ions fingerprint for detailed structural elucidation of the molecule. The abundant fragment ion detected as a singly charged ion at *m/z* 483.07, assigned to Hex-HexNAc-Thr, corresponds to the loss of both sialic acid moieties from the precursor ion. The stripping of only one sialic acid moiety is indicated by the doubly charged ion at *m/z* 386.56 and the singly charged at *m/z* 774.13 along with the ions obtained by further neutral loss of H_2O at *m/z* 756.10 and of CO_2 at *m/z* 730.13. The trisaccharide ion $C_3/B_{1\beta}$ at *m/z* 673.06 is assigned to NeuAc-Hex-HexNAc. Of particular interest for the localization of sialic acid residue are the singly charged ions at *m/z* 470.05, assigned to NeuAc-Hex ($C_{2\alpha}$)

and at *m/z* 493.07, corresponding to NeuAc-HexNAc ($C_3/C_{2\alpha}$). These ions indicating the attachment of the NeuAc moiety represent valuable diagnostic elements for the sialylation pattern of the core 1 *O*-GalNAc mucine molecules. In addition, the ion generated only by the glycosidic-amino acid bond cleavage, with retention of sialic acid entities, is evidenced by the doubly charged ion at *m/z* 481.60, assigned to (NeuAc2)Hex-HexNAc (C_3). In the spectrum in Figure 55, two relevant ring cleavage ions at *m/z* 350.04 and at *m/z* 306.08 corresponding to $^{0,4}A_2$ and $^{0,4}A_2$-CO_2 are detected. According to our previous studies [45, 46] these cleavages are diagnostic for the NeuAcα2-6 linkage.

Characterisation of GT1 ganglioside fraction from normal adult human cerebrum

Gangliosides are sialylated glycosphingolipids (GSLs) consisting of mono- to poly sialylated oligosaccharide chains of variable length attached to the ceramide portion (see page 134). For analysis of complex ganglioside mixtures from biological sources by MS methods, either with ESI or MALDI, specific conditions related to ionisation process and a reliable assessment of the heterogeneity in composition, followed by identification of the molecule architecture, are required.

Optimization of a procedure for capillary-based (-)nanoESI-QTOF MS and tandem MS for characterization of native ganglioside mixture from brain tissues was recently reported [37, 47, 48]. In these studies, requirements for elevated values of ionisation parameters and an extended time for signal acquisition implying relatively high sample consumption for identification and structural elucidation were defined.

To test a feasibility and advantages of the microchip ESI-QTOF MS analysis concerning information that could be provided by both MS and MS/MS, as well as to define corresponding appropriate conditions for the GSL molecular class detection and structural characterization, a rather structurally complex polysialylated ganglioside fraction, GT1, was

chosen as the test sample. The analysed GT1 ganglioside fraction, showing migration properties of GT1b species in high performance thin-layer chromatography (HPTLC), was isolated from the total native ganglioside mixture purified from adult human cerebrum.

For (-) microchip ESI QTOF MS analysis, GT1 sample was dissolved in pure MeOH to a concentration of 5 pmol/µl and an aliquot of 10 µl was loaded into the microchip reservoir. To optimise the ionic current value, the (-)ESI voltage and cone potential were increased from zero following a ramping procedure. The maximum current value was obtained at 3kV ESI and 100 V applied on the sampling cone. In Figure 56a the total ion current (TIC) of GT1 acquired for 6 min is presented.

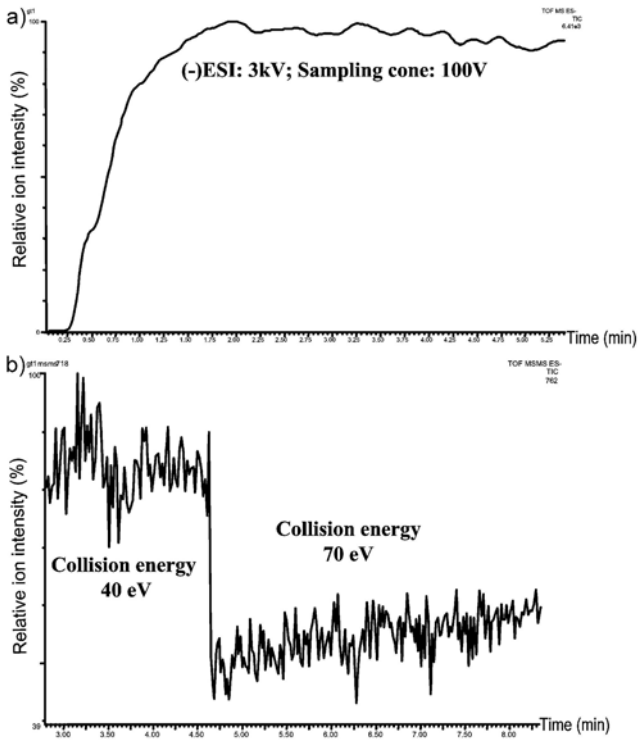


Figure 56. (-) Microchip ESI QTOF TIC of the GT1 ganglioside fraction from normal adult human brain. Maximal ionic current at 3 kV ESI voltage and 100 V sampling cone potential. Solvent : MeOH; Average sample concentration 5 pmol/µl; a) MS mode; b) MS/MS mode for the triply charged ion at m/z 717.50.

The TIC profile indicates a sustained spray varying only by ESI source parameter variation, as well as an efficient ionisation at 3kV ESI and 100 V sampling cone voltages,

proved by the high intensity of ionic current. Starting at 1.5 minute, the signal was acquired for 4 minutes under 3kV and 100 V, to give a constant ESI signal documented by the flat region of TIC. By combining all TIC scans of constant current intensity, a spectrum of high signal/noise was obtained (Figure 57) however, a fair signal/noise ratio was visible already after 60 scans (30 seconds) of acquisition.

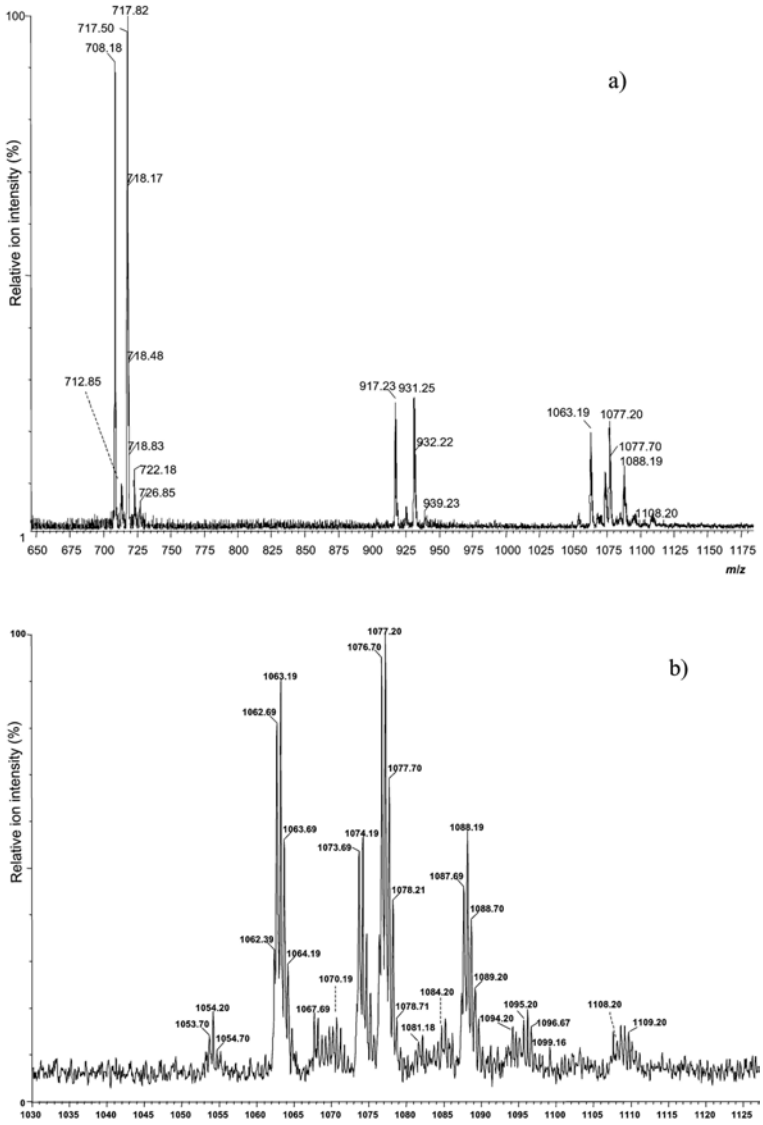


Figure 57. (-) Microchip ESI QTOF MS of the GT1 fraction derived by combining the TIC MS scans at 3kV ESI voltage and 100 V sampling cone potential. a) m/z range: (650-1175); b) zoom out of the m/z range: (1030-1125).

A reproducible compositional mapping of molecular components in the GT1 fraction mixture was obtained from both sets of formed molecular ions, the triply and the doubly charged related to gangliosides containing a number of lipid variants (Table 10). The enlarged area of the spectrum (*m/z* range 1030-1125), containing a rich pattern of doubly charged molecular ions all attributed to GT1 species, is presented in Figure 57b. In particular, each specific GT1 species is characterized by exclusively one triply charged $[M-3H^+]^{3-}$ ion, while the corresponding doubly charged $[M-2H^+]^{2-}$ ion is mostly accompanied by the same charge state $[M+Na^+-3H^+]^{2-}$ and/or $[M-2H^+]^{2-}-H_2O$ ions.

m/z (monoisotopic)	Type of Detected Molecular Ion	Putative Structure
708.18 1062.69 1073.69 1053.70	$[M-3H^+]^{3-}$ $[M-2H^+]^{2-}$ $[M+Na^+-3H^+]^{2-}$ $[M+2H^+]^{2-}-H_2O$	GT1 (d18:1/18:0)
712.85 1070.19 1081.18	$[M-3H^+]^{3-}$ $[M-2H^+]^{2-}$ $[M+Na^+-3H^+]^{2-}$	GT1 (d18:1/19:0)
717.50 1076.70 1087.69 1067.69	$[M-3H]^{3-}$ $[M-2H]^{2-}$ $[M+Na-3H]^{2-}$ $[M+2H^+]^{2-}-H_2O$	GT1 (d18:1/20:0)
1094.20^{L}	$[M+Na^+-3H^+]^{2-}$	GT1 (d18:1/21:1)
722.18 1084.20 1095.20	$[M-3H^+]^{3-}$ $[M-2H]^{2-}$ $[M+Na^+-3H^+]^{2-}$	GT1 (d18:1/21:0)
726.85	$[M-3H^+]^{3-}$	GT1 (d18:1/22:0)
1108.20^{L}	$[M+Na^+-3H^+]^{2-}$	GT1 (d18:1/23:1)
1109.21^{L}	$[M+Na^+-3H^+]^{2-}$	GT1 (d18:1/23:0)

Table 10. Compositional mapping of the purified native GT1 ganglioside fraction (exhibiting HPTLC migration properties of the GT1b species) separated from total ganglioside mixture isolated from adult human brain tissue as detected by (-) microchip-ESI-QTOF MS. d=dihydroxy sphingoid base. ^{L} *low intensity ions.*

The presence of three minor, trace abundant, GT1 molecular species is documented only by very low abundant ions at *m/z* 1094.20, 1108.20 and 1109.21, which can be

assigned by calculation to GT1 (d18:1/21:1), GT1 (d18:1/23:1) and GT1 (d18:1/23:0), respectively.

As obvious from the spectrum pattern, only a low level of *in-source* fragmentation occurs during the MS ionization process, according to the ions assigned to disialylated product ions resulting from the cleavage of one sialic acid residue from the intact trisialylated GT1 species. The NeuAc residue linked to the terminal Gal residue as a monosialo group is known to be more susceptible even to the enzymatic cleavage, such as that by *V. cholerae* neuraminidase. The loss of the NeuAc residue(s) bound to the inner Gal would produce a high extent of the disialo ions along with NeuAc-Gg_4Cer (GM1-like) fragment ions, which are not detectable in the spectrum. A significantly low extent of *in-source* induced fragmentation using the chip-based ESI *vs*. conventional capillary-based ion source was already observed and documented [49], representing one of the major advantages of the chip usage in MS analysis.

It is well-known that in the case of ganglioside fragmentation by CID MS/MS, a long signal acquisition time under variable collision energy is necessary to obtain sufficient abundance of fragment ions diagnostic for the structure elucidation. This approach though successfully applied by using capillary-based ESI [37, 47, 48] often encounters difficulties associated to the spray instability and/or signal interruptions. In case of microchip ESI-MS/MS the triply charged ion detected at *m/z* 717.50 could be successfully fragmented. The TIC profile of the sequencing events given in Figure 56b documents the stability of the MS/MS signal for each of the collision energy values used for this experiment and the corresponding dependence of the ionic current on this parameter. The MS/MS product ion spectrum of the triply charged precursor ion at *m/z* 717.63, corresponding to the GT1 (d18:1/20:0) species is depicted in Figure 58.

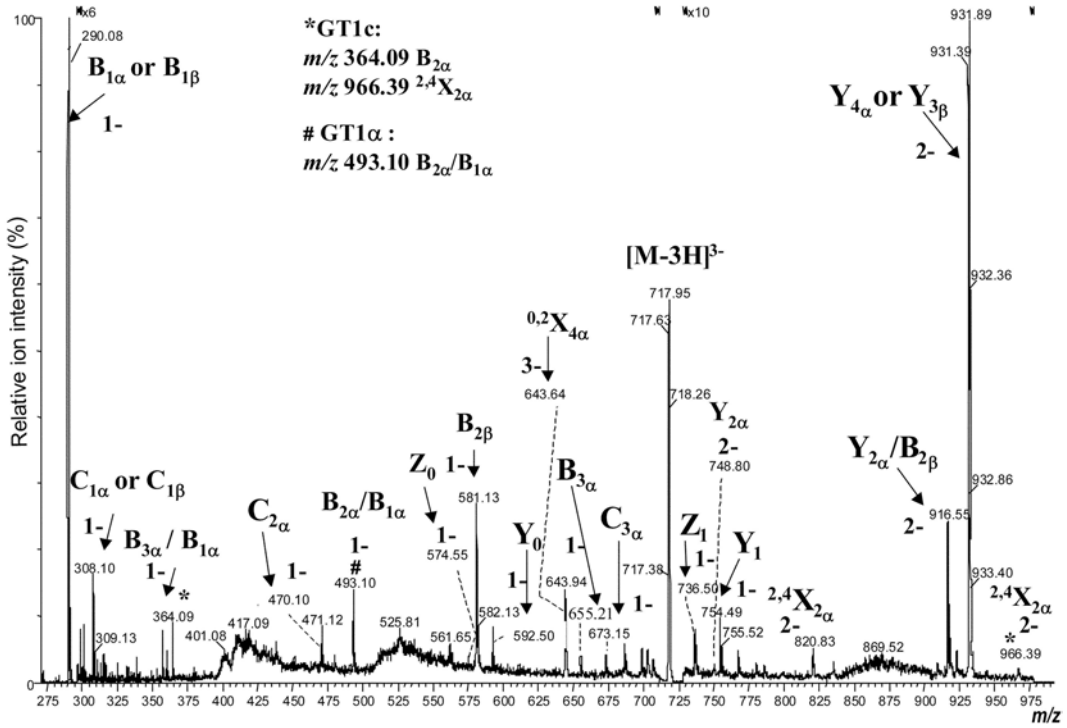


Figure 58. (-) Microchip ESI QTOF MS/MS of the triply charged ion at m/z 717.50 corresponding to GT1 (d18:1/20:0). ESI voltage 3kV. Sampling cone potential 100 V. The spectrum is derived by combining the scans corresponding to 40 and 70 eV in the TIC MS/MS from Figure 56b. The nomenclature for assignment of fragment ions is according to published recommendations [35].

The spectrum was acquired under variable energy (VE-CID) [50] by combining scans at collision energy values of 40 and 70 eV. The fragmentation process gave rise to product ions for reliable assignment of the entire carbohydrate sequence of the GT1b isomer species, as well as to characterize the type of its ceramide portion. However, certain ions indicative for positional isomers, particularly those of GT1a, the GT1c and/or even the GT1a, in the mixture with GT1b, are present as minor components. The ion assignment in the shown spectrum defines the GT1b product ions as represented in the accompanied fragmentation scheme (Figure 59).

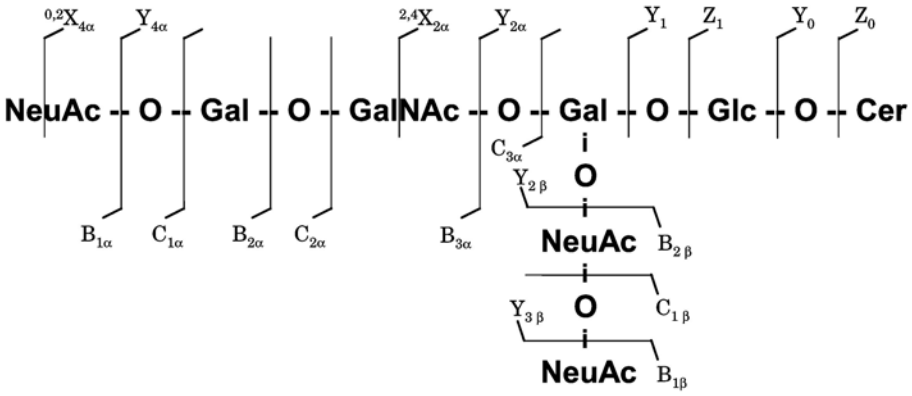


Figure 59. Structure and the corresponding fragmentation scheme of the GT1b species. The product ions detected from the spectrum depicted in Figure 58 are assigned.

The triply charged, still high abundant precursor ion is clearly detectable in the spectrum. As a general characteristic, a predominance of the singly over the doubly and triply charged fragment ions can be pointed out. Also, no sodiated counterparts are observed. The complete sequence of the GT1b is deducible from the ions arising from the non-reducing end of the molecule where dominating Y type ions are accompanied by Z and ring cleavage X type ions appearing at lower abundance, as deduced from the corresponding fragmentation scheme (Figure 59). The d18:1/18:0 type of ceramide is represented by both Y_0 ion and a minor Z_0 at *m/z* 592.50 and 574.55, respectively. The Glc-Cer sequence is also shown by Y_0 and Z_0 counterparts at *m/z* 754.49 and 736.55, respectively. The $Y_{2\alpha}/B_{2\beta}$ ion at *m/z* 916.55 gives rise to the Gal-Glc-Cer sequence which might originate from either GT1a or GT1b isomer after the respective desialylation of either NeuAc or $NeuAc_2$ residue from the inner Gal residue, or might hypothetically originate from an isomer containing the non-substituted inner Gal (potential GT1d isomer). The $^{2,4}X_{2\alpha}$ ion resulting from the ring cleavage at the GalNAc residue corresponds to the inner Gal disialylated structure, GalNAc-($NeuAc_2$)Gal-Glc-Cer, consistent with the GT1b isomer. The disialylation at the inner Gal residue is finally confirmed by the doubly charged $Y_{2\alpha}$ ion at *m/z* 748.80, detected at lower abundance. The highly abundant doubly charged signal at *m/z*

931.39 correspondingly to GT1b isomer, can be attributed to both $Y_{4\alpha}$ and $Y_{3\beta}$ ions giving rise to the presence of disialylated Gg_4Cer sequence; that ion, however, do not indicate the attachment sites of sialic acid residues at the neutral Gg_4 core. Except for the $Y_{2\alpha}$ and the $^{2,4}X_{2\alpha}$ specific for the GT1b isomer, all ions could also arise also from other isomers, such as GT1α, GT1a, GT1c or even, yet not reliably proven to exist, GT1d isomer. The GT1 corresponding sequence was finally corroborated by the triply charged signal at *m/z* 643.64 consistent with 0,2 type ring cleavage at one of the terminal sialic acid residues which, in the case of GT1b species, can be formed from both $^{0,2}X_{4\alpha} - 2H^+$ and $^{0,2}X_{3\beta} - 2H^+$ ion.

Further support and deeper insight into the detailed structural composition of the chosen molecular ion is provided by a set of ions formed from the reducing end. In particular, informative C and B type ions are present and observed to be similarly favoured under the defined conditions. Structurally significant and/or considerably abundant A type ring cleavage ions were not found. The B type ion with the highest intensity at *m/z* 290.08, attributable to the both $B_{1\alpha}$ and $B_{1\beta}$ linkage cleavages, and the accompanied C type counterpart ion, specifying the both $C_{1\alpha}$ and $C_{1\beta}$ cleavages, characterise the cleaved monosialo residue. The presence of the disialo group is evidenced by the abundant $B_{2\beta}$ ion at *m/z* 581.13. The $C_{2\alpha}$ ion (*m/z* 470.12) representing the NeuAc-Gal sequence confirms the attachment of the monosialo group to the terminal Gal residue, which is consistent with the GT1b but also the GT1aα structural motif. The signal might potentially include the ions formed from the sialylated inner Gal residue by multiple internal cleavages. The counterpart signals at *m/z* 673.15 and 655.21, arising from GT1b as $C_{3\alpha}$ and $B_{3\alpha}$ ions, respectively, represent both NeuAc-Gal-GalNAc and Gal-GalNAc-NeuAc sequences, and could therefore arise from any of GT1 isoform. The signal at *m/z* 364.09 corresponding to the Gal-GalNAc sequence, assigned to the GT1b internal cleavage $B_{3\alpha}/B_{1\alpha}$ ion, is however more

characteristic for the GT1c isomer (the $B_{2\alpha}$ ion) bearing the non-substituted terminal Gal-GalNAc residue. The GT1c isoform is further supported by the structure-specific ion at *m/z* 966.39 assigned to $^{2,4}X_{2\alpha}$ ring cleavage ion containing the three sialo structural motif. Finally, the prominent ion at *m/z* 493.10 attributed to the B type NeuAc-GalNAc- ion can arise exclusively from the GT1α isoforms (potentially GT1aα and/or GT1bα), characterised by the sialylation of the GalNAc residue. This ion strongly supports the hypothesised presence of at least one GT1α isoform in the purified GT1 mixture from human brain, containing the GT1b isomer as the major component.

6. FT-ICR analysis of O-glycopeptides

6.1. Material and methods

Chemicals

Analytical-grade methanol was purchased from AppliChem (Darmstadt, Germany) and used without further purification. Aqueous sample solutions were dried in a SpeedVac SPD 111V system from Savant (Düsseldorf, Germany). Prior to chip MS analysis the sample/methanol (MeOH) solutions were centrifuged for one hour in an Eppendorf 5415 C centrifuge (Eppendorf, Hamburg, Germany).

Biological samples

The first sample investigated in this work is a fraction of *O*-glycopeptides extracted from urine of healthy individual. The sample is further denoted Ty2. The second sample is a fraction of *O*-glycopeptides, obtained from urine of the child G.P. suffering from Schindler disease type I, the form of the disease with infantile onset [17, 36]. This mixture is further denoted Gy2. The patient and healthy control subject were age-matched and the samples

were prepared identically. The preparation procedure is in detail described elsewhere [51, 52]. Briefly, for isolation of components, the urine has been filtered, submitted to gel filtration chromatography on Biogel P2, fractionated and desalted by gel filtration chromatography performed on Fractogel TSK HW 50 in 0.01 M pyridinium acetate pH 5.4 as eluting buffer.

Stock sample/methanol (MeOH) solutions of 1µg/µl have been prepared and stored at -70°C. The working aliquots of 20, 12, 7 and 6 pmol/µl (calculated for 2500 Da average molecular mass) were prepared by drying the stock solution and further dissolved in methanol to the appropriate concentration.

Microchip system

Microchips used in this section are exactly the same as those described above (see the paragraph "Microchip system" on page 148). For sample loading, the chip was sandwiched in a home-made chip holder with an integrated reservoir (Figure 60). The chip was positioned into the holder with the microchannel in contact with the reservoir and the front part extruding a few mm. The chip was grounded via a conductive wire connected to the terminal electrode.

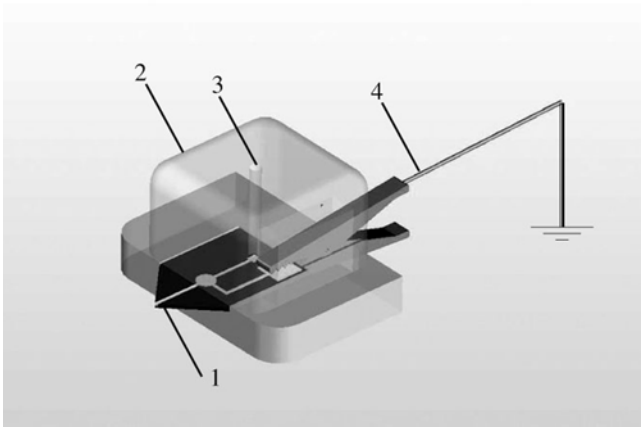


Figure 60. Schematic of the polymer chip holder with integrated reservoir. 1. Polymer chip; 2. Sandwich chip holder; 3. Sample reservoir; 4. Conductive wire.

Between sample analysis, the chip has been flushed with 100 µl methanol to prevent any cross-contamination or carry-over processes. A subsequent (-)chipESI/FT-ICR MS analysis of fresh methanol solution, after flushing the chip, indicated no cross-contamination and/or carry-over processes. Moreover, no retention of the sample species on the inner channel walls was observed. The integrated reservoir of the sandwiched polymer chip turned out to prevent the fast evaporation of the sample solution during the MS experiments and allowed recovering of the unused sample/solution.

Mass spectrometry

Measurements were performed on a Bruker Apex II Fourier Transform Ion Cyclotron Resonance mass spectrometer (FT-ICR MS) (Bruker Daltonik, Bremen, Germany) equipped with a 9.4 T superconducting actively shielded magnet (Magnex Scientific Ltd., Oxford, UK) and a Bruker Apex II Apollo ESI ion source. The chip was coupled to the Bruker Apex II ion source by an in-laboratory constructed mounting system. The interface consists of a metal plate mounted to the Apollo ion source by two 90° brackets. The plate and the 90° brackets featured two slots for the screws, thus providing x- and z-axis movement and some y-axis variability. The chip holder was attached to the metal plate and carefully positioned toward the entrance orifice of the Apex II capillary. Mounting of the chip interface system did not claim for any definitive dismantling or irreversible mechanical modifications to either of the original Appolo source components. Moreover, the exchange between the original source and chip interface is quick and simple.

The Apex II metal-coated glass capillary was kept in the range of 1500 – 2500 V while the chip was grounded. The capillary exit voltage was set to –79 V to reduce the *in-source* ion decomposition to a minimum. The generated ions were accumulated for 0.05 to 0.3 s in the hexapole located after the second skimmer of the ion source and then transferred

into the ion cyclotron resonance cell. The ions were trapped by standard trapping procedures.

All mass spectra were externally calibrated and acquired in the broad band mode with 1 Mpoints/scan. The calculation of the theoretical masses displayed in Table 11 and Table 12 was carried out by using the values for atomic weights taken from De Laeter et al. [53] and including the electron mass (cf. Mamer et al. [54]).

6.2. Results and discussion

The polymer chip was assembled into the sandwich chip holder and fixed onto the in-laboratory constructed mounting system. The mounting system was conceived and designed to provide a robust and viable interfacing of the polymer chip to the FT-ICR MS. The positioning/alignment of the polymer chip on the *x*-, *y*- and *z*- axis turned out to be a crucial step in the initiation and long-term maintaining of the electrospray, in particular with respect to the direct-spray configuration of the Bruker Apex II instrument.

Ty2 mixture of *O*-GalNAc glycosylated sialylated peptides was dissolved to a concentration of approximately 7 pmol/µl in methanol and 20 µl of sample/methanol solution were loaded into the incorporated reservoir. The negative ion mode electrospray process could be initiated at 1500 V applied to the Apex II capillary and –79 V potential of the capillary exit. Under these well-defined conditions a constant and stable spray accompanied by a high ionic intensity was generated. The signal was acquired over 155 scans at a flow rate of about 200 nl/min. The resulted spectrum is depicted in Figure 61.

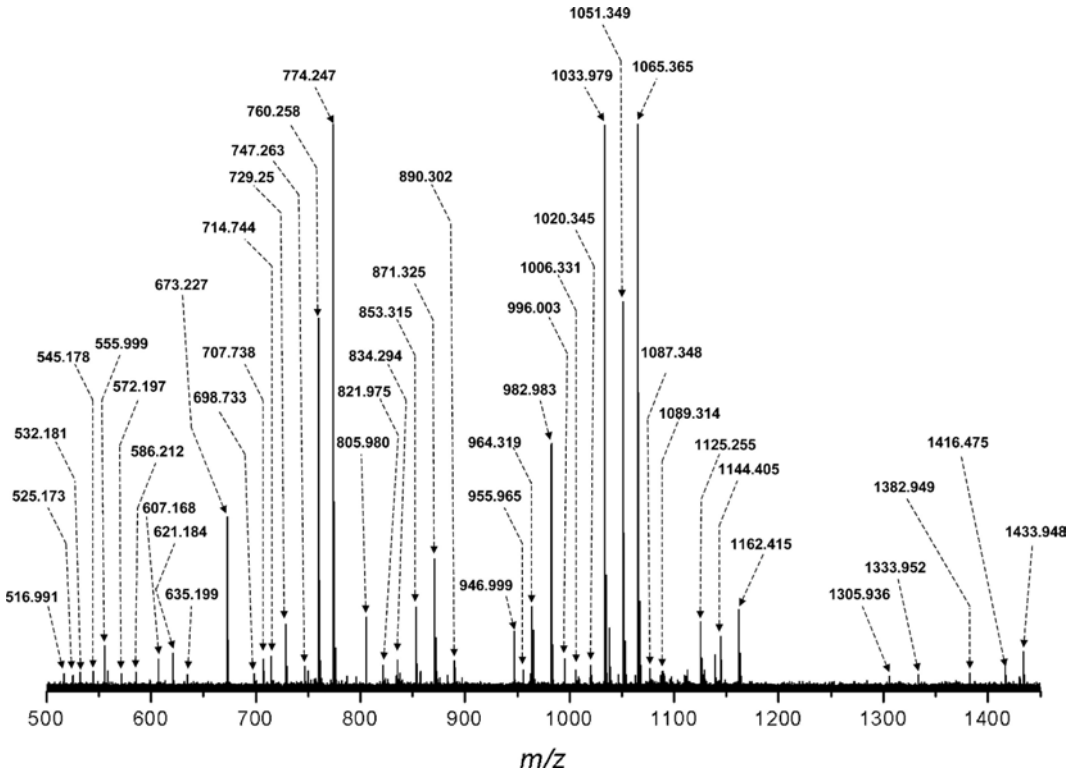


Figure 61. (-)ChipESI/FT-ICR MS of the Ty2 mixture of O-glycosylated peptides from urine of healthy individual. Sample concentration: 7 pmol/μl in methanol. Number of scans: 155.

The spectrum shows a high heterogeneity of the mixture regarding the number of species and structural variety, represented in the spectrum by a large number of singly and doubly charged ions of high signal-to-noise ratio. The most abundant species assigned in Ty2 mixture are disialylated saccharides linked to Ser and Thr. The Ser- linked saccharides are documented by the singly charged ion at *m/z* 1051.349 and its corresponding doubly charged ion at *m/z* 525.173 assigned to Neu5Ac2HexHexNAc-Ser. The doubly charged ion at *m/z* 707.738 is assigned to Neu5Ac2Hex2HexNAc2-Ser and the doubly charged ion at *m/z* 890.302 to Neu5Ac2Hex3HexNAc3-Ser, while the Ser- linked monosialylated trisaccharide is visible as a singly charged ion at *m/z* 760.258. Their homologous Thr-linked structures are detected at *m/z* values shifted by the corresponding difference between Ser and Thr, according to their respective charge states. Two free oligosaccharides are evidenced as singly charged ions at *m/z* 673.227 corresponding to Neu5AcHexHexNAc and

at *m/z* 964.319 attributed to Neu5Ac2HexHexNAc, respectively. An interesting aspect arising from the spectrum in Figure 61 is the presence of a number of dehydrated sialylated glycopeptides, which could originate either from an *in-source* neutral loss of water or as already metabolically generated components present in the mixture. For instance, the dehydrated Neu5Ac2Hex2HexNAc2-Ser is detected as a doubly charged ion at *m/z* 698.733, Neu5Ac2HexHexNAc-Thr as a singly charged ion at *m/z* 1047.355, Neu5Ac2HexHexNAc-Thr-Pro as a singly charged ion at *m/z* 1144.416 and Neu5AcHexHexNAc-Thr-Pro as a singly charged ion at *m/z* 853.315. It should be pointed out, that the abundance of the disialylated saccharides is higher than that of the monosialylated species. This aspect demonstrates the feasibility of the optimized (-) chipESI/FT-ICR MS approach to exhibit only a very low degree of *in-source* decay related to the labile sialic acid moiety, therefore to consequently fulfil expectations defined as essential for detection of intact sialylated glycoconjugates. All ions detected and assigned in the Ty2 mixture under the assumption of a mass accuracy below 10 ppm are listed in Table 11.

m/z		Charge state	Proposed structure
Exp.	Theor.		
525.173	527.175	2	$Neu5Ac_2HexHexNAc$-Ser
532.181	532.183	2	$Neu5Ac_2HexHexNAc$-Thr
673.227	673.230	1	Neu5AcHexHexNAc
698.733	698.736	2	$Neu5Ac_2Hex_2HexNAc_2$-Ser-H_2O
707.738	707.741	2	$Neu5Ac_2Hex_2HexNAc_2$-Ser
714.744	714.749	2	$Neu5Ac_2Hex_2HexNAc_2$-Thr
760.258	760.262	1	Neu5AcHexHexNAc-Ser
774.278	774.278	1	Neu5AcHexHexNAc-Thr
853.315	853.320	1	Neu5AcHexHexNAc-Thr-Pro-H_2O
871.325	871.331	1	Neu5AcHexHexNAc-Thr-Pro
890.302	890.307	2	$Neu5Ac_2Hex_3HexNAc_3$-Ser
964.319	964.326	1	$Neu5Ac_2HexHexNAc$
1047.355	1047.363	1	$Neu5Ac_2HexHexNAc$-Thr-H_2O
1051.349	1051.358	1	$Neu5Ac_2HexHexNAc$-Ser
1065.365	1065.374	1	$Neu5Ac_2HexHexNAc$-Thr
1087.348	1087.356	1	$Neu5Ac_2HexHexNAc$-Thr (Na)
1139.401	1139.410	1	$Neu5AcHex_2HexNAc_2$-Thr

1144.405	1144.416	1	$Neu5Ac_2HexHexNAc\text{-}Thr\text{-}Pro\text{-}H_2O$
1162.415	1162.426	1	$Neu5Ac_2HexHexNAc\text{-}Thr\text{-}Pro$
1416.475	1416.490	1	$Neu5Ac_2Hex_2HexNAc_2\text{-}Ser$

Table 11. Assignment of the ions detecterd in the Ty2 mixture by microchip ESI-FT-ICR-MS in the negative ion mode.

The advantage of the (-) chipESI/FT-ICR MS approach to combine an efficient ionization and ultrahigh resolution and mass accuracy of detection is clearly highlighted here by the high number, 20, of ionic species (Table 11) unambiguously assigned within an average mass accuracy of 5.89 ppm.

An extensive study regarding the sensitivity and concurrently an efficient production of structural data has been carried out by (-)chipESI/FT-ICR MS on the Gy2 mixture of glycopeptides. For comparative studies exploring the limits of sensitivity exhibited by the novel (-)chipESI/FT-ICR MS method, at which the level of mixture complexity can still be reliably assessed, the Gy2 mixture has been screened at different sample solution concentrations. Thus, the (-)chipESI/FT-ICR MS experiments carried out on the Gy2 mixture at approximately 20 pmol/µl in methanol gave raise to the spectrum in Figure 62. By applying the same optimised conditions for ionisation/detection by (-)chip/ESI-FT-ICR MS a highly stable and constant spray as well as a high signal-to-noise ratio was obtained after 50 scans signal recording (Figure 62). A number of 27 ionic species could be unambiguously identified and assigned within an average mass accuracy of 5.41 ppm.

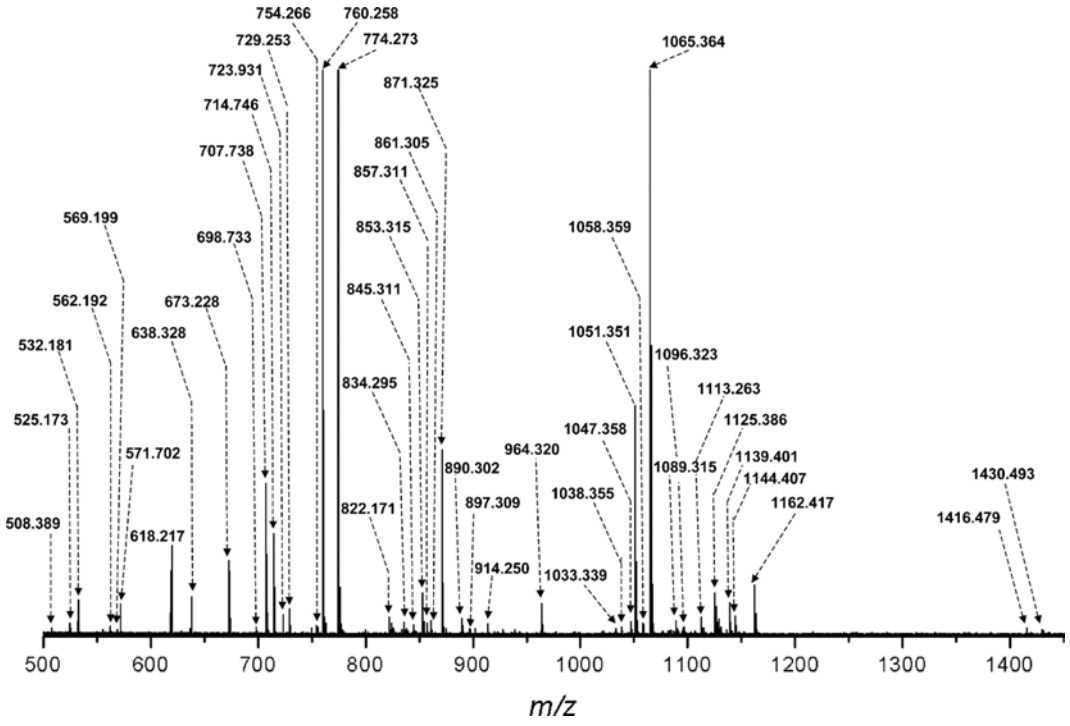


Figure 62. (-)chipESI-FT-ICR-MS of the Gy2 mixture of O-glycosylated peptides from urine of a patient diagnosed with Schindler disease. Sample concentration: 20 pmol/µl in methanol. Number of scans: 50.

In comparison with the spectrum of the Ty2 mixture, a higher content of pentasaccharides is present as evidenced by the doubly charged ion at *m/z* 562.193 and accordingly by a singly charged ion at *m/z* 1125.386 corresponding to Neu5AcHex2HexNAc2-Ser. The doubly charged ion at *m/z* 569.199, along with the singly charged one at *m/z* 1139.401 are assigned to a homologue structure of Neu5AcHex2HexNAc2-Thr. Additionally, the hexasaccharides linked to Ser and Thr bearing two sialic acid moieties are visible here in a higher abundance than in Ty2, as doubly charged ions at *m/z* 707.738 and 714.746 and as singly charged ions at *m/z* 1416.479 and 1430.493, respectively. A nonasaccharide bearing three sialic acid moieties is detected at a fair abundance as a doubly charged ion at *m/z* 1058.354 and assigned to a sodiated dehydrated Neu5Ac3Hex2HexNAc4-Ser with a mass accuracy of 4.6 ppm. The ions

identified in the Gy2 mixture and assigned with a mass accuracy below 10 ppm are listed in Table 12.

m/z		Charge state	Proposed structure
Exp.	Theor.		
525.173	527.175	2	Neu5Ac2HexHexNAc-Ser
532.181	532.183	2	Neu5Ac2HexHexNAc-Thr
562.192	562.193	2	Neu5AcHex2HexNAc2-Ser
569.199	569.201	2	Neu5AcHex2HexNAc2-Thr
571.702	571.704	2	Neu5Ac2HexHexNAc-Thr-Pro-H2O
673.228	673.230	1	Neu5AcHexHexNAc
698.733	698.736	2	Neu5Ac2Hex2HexNAc2-Ser-H2O
707.738	707.741	2	Neu5Ac2Hex2HexNAc2-Ser
714.746	714.749	2	Neu5Ac2Hex2HexNAc2-Thr
760.258	760.262	1	Neu5AcHexHexNAc-Ser
774.273	774.278	1	Neu5AcHexHexNAc-Thr
853.315	853.320	1	Neu5AcHexHexNAc-Thr-Pro-H2O
871.325	871.331	1	Neu5AcHexHexNAc-Thr-Pro
890.302	890.307	2	Neu5Ac2Hex3HexNAc3-Ser
897.309	897.315	2	Neu5Ac2Hex3HexNAc3-Thr
964.320	964.326	1	Neu5Ac2HexHexNAc
1033.339	1033.347	1	Neu5Ac2HexHexNAc-Ser-H2O
1047.358	1047.363	1	Neu5Ac2HexHexNAc-Thr-H2O
1051.351	1051.358	1	Neu5Ac2HexHexNAc-Ser
1058.351	1058.354	2	Neu5Ac3Hex2HexNAc4-Ser-H2O (Na)
1065.364	1065.374	1	Neu5Ac2HexHexNAc-Thr
1125.386	1125.395	1	Neu5AcHex2HexNAc2-Ser
1139.401	1139.410	1	Neu5AcHex2HexNAc2-Thr
1144.407	1144.416	1	Neu5Ac2HexHexNAc-Thr-Pro-H2O
1162.417	1162.426	1	Neu5Ac2HexHexNAc-Thr-Pro
1416.479	1416.490	1	Neu5Ac2Hex2HexNAc2-Ser
1430.493	1430.506	1	Neu5Ac2Hex2HexNAc2-Thr

Table 12. Assignment of the ions detected in the Gy2 mixture by microchip ESI-FT-ICR-MS in the negative ion mode.

Even though, under the optimised conditions for glycopeptide ion formation and detection, a fair coverage of ionic species assignment could be achieved. A large number of structures represented in the spectrum by doubly and singly charged ions at lower abundance could be assigned only at a mass accuracy above 10 ppm. This aspect can be rationalised by two main factors: a) space charge effects, which can occur for high number/type of ions present in narrow *m/z* ranges [55-57] and b) overlapping of formally

isobaric structures which can be frequently encountered when dealing with highly complex carbohydrate mixture originating from biological sources [39].

Further, the Gy2 mixture was stepwise diluted in methanol down to a concentration of 3 pmol/µl and subjected to (-)chipESI/FT-ICR MS investigation. The (-)chipESI/FT-ICR MS of the Gy2 mixture at 12 and 6 pmol/µl are depicted in Figure 63.

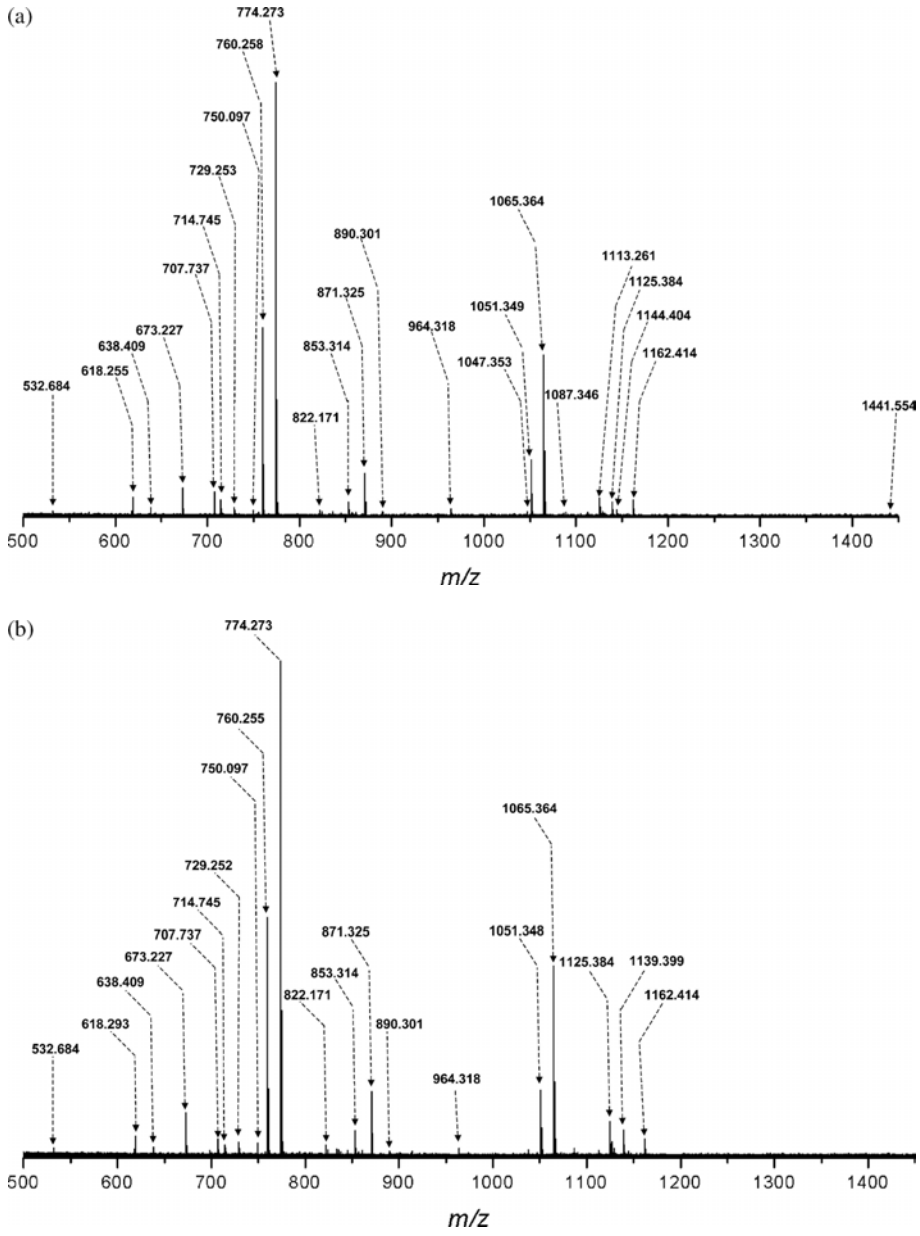


Figure 63. (-)ChipESI FT-ICR-MS of the Gy2 mixture of O-glycosylated peptides from urine of a patient diagnosed with Schindler disease. (a) sample concentration: 12 pmol/µl. Number of scans: 50. (b) sample concentration: 6 pmol/µl. Number of scans: 100.

Under these sample dilution conditions a good coverage of glycopeptide ion species at a fair signal-to-noise ratio to render structural data could still be obtained. No significant difference with respect to ionic species detection is observed between the spectra in Figure 63a) and b), thereby demonstrating the potential of the novel (-)chipESI/FT-ICR MS to exhibit high sensitivity along with high mass accuracy and resolution of detection. Moreover, the long-lasting and constant spray generated from the polymer chip enabled the MS signal acquisition for an extended time period, resulting in a high signal-to-noise ratio. This aspect is more clearly illustrated by the spectrum in Figure 63b) where 100 scans were acquired using a sample concentration of 6 pmol/µl and a flow rate of approximately 200 nl/min. A significant peculiarity of the polymer microchip coupled to FT-ICR MS is the capability to exhibit a spray of high stability over a long time frame of signal acquisition using biological samples. In the case of capillary based nanospray analysis of similarly prepared glycopeptide fractions [18] using in-house pulled glass capillaries, besides the poor reproducibility of the sprayer shape, often blockages of the tip inducing signal interruption were noticed. A visual inspection showed clogging of the tip by salt deposition. The polymer chip turned out to be more tolerant to salt deposition and insoluble impurities inevitably present in biological samples after laborious preparation stages. During the entire analysis time, the microchip did not get blocked by any salt crystals or impurities. This can be most probably attributed to the microsprayer dimensions and higher flow rate exhibited by the polymer chip.

7. Conclusion

The primary goal of this study was to assess the potential of microfabricated polymer microsprayers for the analysis of carbohydrates. The requirements were: a spray

stable enough to acquire signals over tens of spectra and perform tandem mass spectrometry, a sensitivity comparable to other spraying techniques, and limited *in-source* fragmentation of labile moieties such as sialic acids. First, the coupling of the polymer microsprayers was readily obtained on the Z-spray source of the QoTOF-MS (Micromass, Manchester, UK), and good spray stability was obtained under gravity flow, as shown in Figure 56. The coupling of the microfabricated microsprayer was more difficult with the Appolo source of the FT-ICR-MS (Bruker Daltonics, Bremen, Germany); this is probably due to the direct in-line configuration of the source. In particular, the positioning of the microchip with respect to the heated capillary inlet appeared to be critical. However, the LCQ Duo (Thermo Electron, San Jose, CA, USA) used in our laboratory employs the same kind of configuration without the same difficulties. Nevertheless, once the electrospray was started, it could be maintained as long as solution was provided to the microchip reservoir.

The concentrations used in this study and in others from the group of Jasna Peter-Katalinic on comparable samples are reported below in Table 13:

Electrospray technique	Sample	Concentration (pmol/µl)	Instrument	Reference
Polymer microchip	Ty (glycopeptides from healthy urine)	1.5-5	QTOF	[58]
Polymer microchip	Ty2 (glycopeptides from healthy urine)	7	FT-ICR	[59]
Polymer microchip	Gy2 (glycopeptides from urine of a Schindler patient)	6-20	FT-ICR	[59]
Advion ESI-Chip	Q5 fraction of urine from Schindler patient	1-7	FT-ICR	[60]
Advion ESI-Chip	Ty (glycopeptides from healthy urine)	5	FT-ICR	[60]

Glass nanospray	BPY1 (glycopeptides of urine from Schindler patient)	5	FT-ICR	[18]
Advion ESI-Chip	BQ5 (glycopeptides of urine from Schindler patient)	5	QTOF	[46]
Polymer microchip	GT1 (gangliosides from brain)	5	QTOF	[58]
Advion ESI-Chip	Gangliosides from human brain	2-15	QTOF	[49]
Glass nanospray	gangliosides	0.6-20	QTOF	[61]

Table 13. Concentrations used in this study and others from the group of Pr Peter-Katalinic.

As shown above, the concentrations of glycopeptides and gangliosides analysed from the microfabricated microsprayers are fully comparable to those from Advion ESI-Chips (i.d. of 9 µm) and pulled glass nanospray capillaries, which are known as the most sensitive electrospray techniques to date [29, 30, 62]. Unfortunately no side-by-side comparison (same sample, same instrument) could be done, but these results are in agreement with the results reported in Chapter 3 and in [63]. This result is in contradiction with the current know-how in electrospray: the microfabricated microsprayer has large dimensions (120 µm wide, 45 µm deep in a half-moon shape), which is supposed to lead to a worst sensitivity. Moreover, as shown in Figure 53, the electrospray generated from the microchip consists in a large cone generated relatively far from mass spectrometer inlet (typically 5-10 mm), whereas pulled capillary nanosprays are used with distances of 1-5 mm); consequently, the sampling of the electrospray jet is believed to be much better in the case of close spraying with nanosprays than with far-spraying microsprays [29]. No clear explanation could be provided so far for these comparable sensitivities obtained with microchips and nanosprays. One possible hypothesis is that the shape of the microchannel outlet or nozzle influence the size of droplet formed: in two independent studies [64, 65], a

triangular nozzle (possibly with concave walls) has been found to be an optimal nozzle cross section in terms of minimisation of droplet size; whereas these two studies were performed for hydrodynamic spraying, it is probable that the same effect applies for electrospray. The whole rationale is that particular shapes maximise the pressure applied on the droplet emitting zone by the flow rate at constant typical dimension (triangle edge versus circle diameter).

Moreover, it is usually admitted that ionisation is more gentle when nanosprays are used together with low ionisation potential, which is crucial for the analysis of fragile molecules such as oligosaccharides or intact protein complexes [30, 66]. Here, the use of microfabricated microsprayers with relatively high ionisation potential was found to result in qualitatively less *in-source* fragmentation of glycans (usually identified from the presence of high intensity sialic acid ions in the spectra) (Alina Zamfir, personal communication). Though somehow contrary to the commonly admitted know-how in electrospray ionisation, this feature has been also identified independently by the group of Alain Van Dorsselaer in the study of intact protein complexes (roundtable "Non-covalent complexes", 21ème Journées Françaises de Spectrométrie de Masse, Strasbourg, September 14-17, 2004): nanospray cannot be used in routine for the analysis of protein non-covalent complexes because it often leads to disruption of the complex, whereas microspray tends to better preserve the complex. No clear rationale has been provided so far; the only hypothesis that can be put forward is that microsprays are usually used with further Taylor cone/ mass spectrometer inlet distances that nanosprays; it is thus possible that when microsprays are used, ions experience significant collisional cooling with ambient gas before entering the low pressure compartment of the mass spectrometer. This hypothesis still has to be tested by measuring ions internal energies [67] by the ion survivor method [68] and comparing them between micro- and nanosprays.

Basically, the methodology systematised by Vekey for ions internal energy measurement is the following (which is an oversimplified description of the process; for more detailed information, the reader is referred to [67, 68]): well characterised molecules are electrosprayed; under optimised conditions, they will not fragment along the ion optics and mass analyser, and their mass spectrum will consist only in their characteristic intact peaks. Now if one acquisition parameter along the ion optics (e.g. the tube lens offset that accelerates or slows down ions before they enter the skimmer-octopole region on the Thermo Electron LCQ Duo) is tuned so that fragmentation occurs, the extent of fragmentation can be measured from the mass spectra to build a breakdown curve, such as the one below:

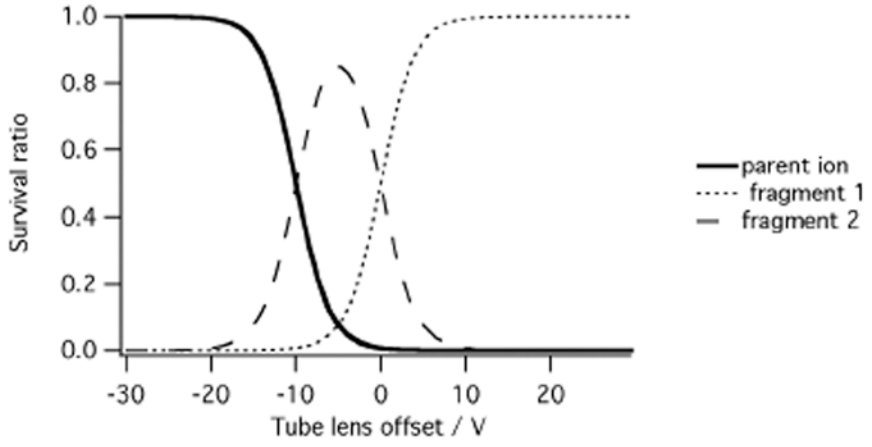


Figure 64. Typical breakdown curve, showing the disappearance of the parent ion and the appearance of two fragments as the tube lens offset increases.

In parallel, molecular kinetic calculation (such as RRKM theory) allows to calculate at which energy different bounds in the parent ion fragment, giving rise to fragments 1 and 2. In a third stage, the mass spectrometric experiment (here the tuning of the tube lens offset) is modelled in order to calculate how much energy is brought to the gas phase ions during the experiment. The difference between the energy brought (in the case here) by tube lens acceleration at breakdown values and the energy necessary to produce the breakdown

of the parent ion and appearance of the fragments is the internal energy that was carried by gas phase ions before experiencing tube lens acceleration. Drahos and Vekey have developed a software called MassKinetics to calculate fragmentation energies of ions, and how much energy is brought to gas phase ions by a given mass spectrometric experiment [68]. This methodology has been applied by e.g. Gabelica *et al* to study the influence of various instrumental parameters on electrosprayed ions energy [69]:

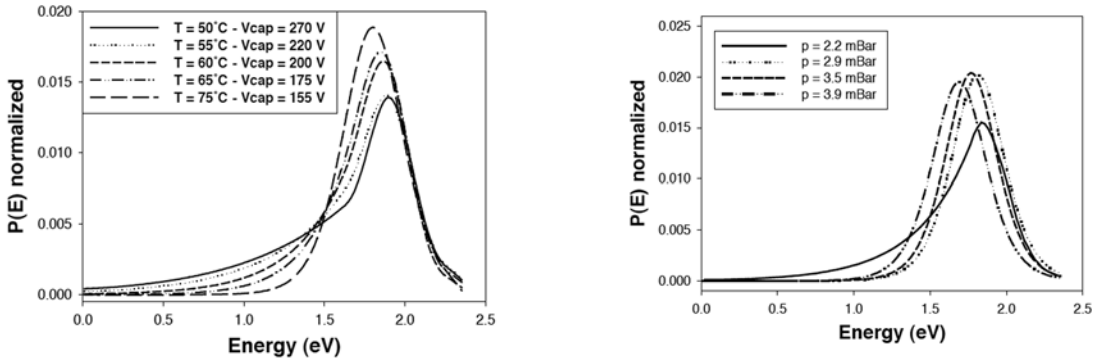


Figure 65. Effect of source temperature and cone voltage on internal energy distributions (left), and of source pressure on internal energy distributions (right). Reprinted with permission from [69].

By following the same approach, it may be possible to compare internal energies of ions generated from a microfabricated microsprayer with those generated by pulled capillary nanospray. A lower internal energy with microfabricated microsprayers would prove that ionisation is softer with microfabricated microsprayers, and would be in line with the observation that oligosaccharides are less fragmented with the microchip compared to comparable techniques.

8. References

[1] Stick, R. V., Carbohydrates: the sweet molecules of life, Academic Press, San Diego London, 2001.

[2] Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G., Marth, J., Essentials of glycobiology, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1999.
[3] Laine, R. A., Information capacity of the carbohydrate code, *Pure Appl. Chem.*, **1997**, 69(9): 1867-1873.
[4] Chester, M. A., Nomenclature of glycolipids, *Pure Appl. Chem.*, **1997**, 69(12): 2475-2487.
[5] Rebois, R. V., Fishman, P. H., Deglycosylated Human Chorionic-Gonadotropin - an Antagonist to Desensitization and Down-Regulation of the Gonadotropin Receptor-Adenylate Cyclase System, *J. Biol. Chem.*, **1983**, 258(21): 2775-2778.
[6] Handel, T. M., Johnson, Z., Crown, S. E., Lau, E. K., Sweeney, M., Proudfoot, A. E., Regulation of protein function by glycosaminoglycans - as examplified by chemokines, *Ann. Rev. Biochem.*, **2005**, 74(1): 385-410.
[7] Morell, A. G., Irvine, R. A., Sternlie.I, Scheinbe.Ih, Ashwell, G., Physical and Chemical Studies on Ceruloplasmin .V. Metabolic Studies on Sialic Acid-Free Ceruloplasmin in Vivo, *J. Biol. Chem.*, **1968**, 243(1): 155-159.
[8] Kansas, G. S., Selectins and their ligands: Current concepts and controversies, *Blood*, **1996**, 88(9): 3259-3287.
[9] Ley, K., Kansas, G. S., Selectins in T-cell recruitment to non-lymphoid tissues and sites of inflammation, *Nat. Rev. Immunol.*, **2004**, 4(5): 325-335.
[10] Dube, D. H., Bertozzi, C. R., Glycans in cancer and inflammation. Potential for therapeutics and diagnostics, *Nat. Rev. Drug Discov.*, **2005**, 4(6): 477-488.
[11] Van Schaftingen, E., Jaeken, J., Phosphomannomutase deficiency is a cause of carbohydrate-deficient glycoprotein type I, *FEBS Lett.*, **1995**, 377(3): 318-320.
[12] Marquardt, T., Denecke, J., Congenital disorders of glycosylation: review of their molecular bases, clinical presentations and specific therapies, *Eur. J. Pediatr.*, **2003**, 162(6): 359-379.
[13] Hanisch, F. G., O-glycosylation of the mucin type, *Biol. Chem.*, **2001**, 382(2): 143-149.
[14] Jaeken, J., Matthijs, G., Congenital disorders of glycosylation, *Annu. Rev. Genomics Hum. Genet.*, **2001**, 2(1): 129-151.
[15] Jaeken, J., Congenital disorders of glycosylation (CDG): It's all in it!, *J. Inherit. Metab. Dis.*, **2003**, 26(2): 99-118.
[16] Garman, S. C., Hannick, L., Zhu, A., Garboczi, D. N., The 1.9 angstrom structure of alpha-N-acetylgalactosaminidase: Molecular basis of glycosidase deficiency diseases, *Structure*, **2002**, 10(3): 425-434.
[17] Vandiggelen, O. P., Schindler, D., Kleijer, W. J., Huijmans, J. M. G., Galjaard, H., Linden, H. U., Peter-Katalinic, J., Egge, H., Dabrowski, U., Cantz, M., Lysosomal alpha-N-acetylgalactosaminidase deficiency - a new inherited metabolic disease, *Lancet*, **1987**, 2(8562): 804.
[18] Froesch, M., Bindila, L., Zamfir, A., Peter-Katalinic, J., Sialylation analysis of O-glycosylated sialylated peptides from urine of patients suffering from Schindler's disease by Fourier transform ion cyclotron resonance mass spectrometry and sustained off-resonance irradiation collision-induced dissociation, *Rapid Commun. Mass Spectrom.*, **2003**, 17(24): 2822-2832.
[19] Meezan, E., Wu, H. C., Black, P. H., Robbins, P. W., Comparative studies on carbohydrate-containing membrane components of normal and virus-transformed mouse fibroblasts. 2. Separation of glycoproteins and glycopeptides by Sephadex chromatography, *Biochemistry*, **1969**, 8(6): 2518-&.

[20] Fredman, P., Mansson, J. E., Dellheden, B., Bostrom, K., von Holst, H., Expression of the GM1-species, NeuN -GM1, in a case of human glioma, *Neurochem. Res.*, **1999**, 24(2): 275-279.

[21] Wagener, R., Rohn, G., Schillinger, G., Schroder, R., Kobbe, B., Ernestus, R. I., Ganglioside profiles in human gliomas: Quantification by microbore high performance liquid chromatography and correlation to histomorphology and grading, *Acta Neurochir.*, **1999**, 141(12): 1339-1345.

[22] Zaia, J., Mass spectrometry of oligosaccharides, *Mass Spectrom. Rev.*, **2004**, 23(3): 161-227.

[23] Rohner, T. C., Lion, N., Girault, H. H., Electrochemical and theoretical aspects of electrospray ionization, *Phys. Chem. Chem. Phys.*, **2004**, 6(12): 3056-3068.

[24] Cech, N. B., Krone, J. R., Enke, C. G., Predicting electrospray response from chromatographic retention time, *Anal. Chem.*, **2001**, 73(2): 208-213.

[25] Bourne, E. J., Stacey, M., Tatlow, J. C., Tedder, J. M., Studies on trifluoroacetic acid. Part 1. Trifluoroacetic anhydride as a promoter of ester formation between hydroxy-compounds and carboxylic acids, *J. Chem. Soc.*, **1949**, Nov(1): 2976-2979.

[26] Ciucanu, I., Kerek, F., A simple and rapid method for the permethylation of carbohydrates, *Carbohydr. Res.*, **1984**, 131(2): 209-217.

[27] Dell, A., Preparation and desorption mass spectrometry of permethyl and peracetyl derivatives of oligosaccharides, *Methods Enzymol.*, **1990**, 193(1): 647-660.

[28] Wilm, M., Mann, M., Analytical properties of the nanoelectrospray ion source, *Anal. Chem.*, **1996**, 68(1): 1-8.

[29] Karas, M., Bahr, U., Dulcks, T., Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine, *Fresenius J. Anal. Chem.*, **2000**, 366(6-7): 669-676.

[30] Bahr, U., Pfenninger, A., Karas, M., Stahl, B., High sensitivity analysis of neutral underivatized oligosaccharides by nanoelectrospray mass spectrometry, *Anal. Chem.*, **1997**, 69(22): 4530-4535.

[31] Doroshenko, V. M., Laiko, V. V., Taranenko, N. I., Berkout, V. D., Lee, H. S., Recent developments in atmospheric pressure MALDI mass spectrometry, *Int. J. Mass Spectrom.*, **2002**, 221(1): 39-58.

[32] Gabelica, V., Schulz, E., Karas, M., Internal energy build-up in matrix-assisted laser desorption/ionization, *J. Mass Spectrom.*, **2004**, 39(6): 579-593.

[33] Konn, D. O., Murrell, J., Despeyroux, D., Gaskell, S. J., Comparison of the effects of ionization mechanism, analyte concentration, and ion "cool-times" on the internal energies of peptide ions produced by electrospray and atmospheric pressure matrix-assisted laser desorption ionization, *J. Am. Soc. Mass Spectrom.*, **2005**, 16(5): 743-751.

[34] O'Connor, P. B., Budnik, B. A., Ivleva, V. B., Kaur, P., Moyer, S. C., Pittman, J. L., Costello, C. E., A high pressure matrix-assisted laser desorption ion source for Fourier transform mass spectrometry designed to accommodate large targets with diverse surfaces, *J. Am. Soc. Mass Spectrom.*, **2004**, 15(1): 128-132.

[35] Domon, B., Costello, C. E., A Systematic Nomenclature for Carbohydrate Fragmentations in Fab-Ms Ms Spectra of Glycoconjugates, *Glycoconjugate J.*, **1988**, 5(4): 397-409.

[36] Linden, H. U., Klein, R. A., Egge, H., Peter-Katalinic, J., Dabrowski, J., Schindler, D., Isolation and structural characterization of sialic acid-containing glycopeptides of the omicron-glycosidic type from the urine of 2 patients with an hereditary-deficiency in alpha-N-acetylgalactosaminidase sctivity, *Biol. Chem. Hoppe-Seyler*, **1989**, 370(7): 661-672.

[37] Metelmann, W., Vukelic, Z., Peter-Katalinic, J., Nano-electrospray ionization time-of-flight mass spectrometry of gangliosides from human brain tissue, *J. Mass Spectrom.*, **2001**, 36(1): 21-29.
[38] Peter-Katalinic, J., Analysis of Glycoconjugates by Fast-Atom-Bombardment Mass-Spectrometry and Related Ms Techniques, *Mass Spectrom. Rev.*, **1994**, 13(1): 77-98.
[39] Zamfir, A., Peter-Katalinic, J., Capillary electrophoresis-mass spectrometry for glycoscreening in biomedical research, *Electrophoresis*, **2004**, 25(13): 1949-1963.
[40] Thanawiroon, C., Rice, K. G., Toida, T., Linhardt, R. J., Liquid chromatography/mass spectrometry sequencing approach for highly sulfated heparin-derived oligosaccharides, *J. Biol. Chem.*, **2004**, 279(4): 2608-2615.
[41] Rossier, J. S., Vollet, C., Carnal, A., Lagger, G., Gobry, V., Girault, H. H., Michel, P., Reymond, F., Plasma etched polymer microelectrochemical systems, *Lab Chip*, **2002**, 2(3): 145-150.
[42] Rossier, J. S., Reymond, F., Michel, P., Polymer microfluidic chips for electrochemical and biochemical analyses, *Electrophoresis*, **2002**, 23(6): 858-867.
[43] Gobry, V., van Oostrum, J., Martinelli, M., Rohner, T., Rossier, J. S., Girault, H. H., Microfabricated polymer injector for direct mass spectrometry coupling, *Proteomics*, **2002**, 2(4): 405-412.
[44] Morier, P., Vollet, C., Michel, P. E., Reymond, F., Rossier, J. S., Gravity-induced convective flow in microfluidic systems: Electrochemical characterization and application to enzyme-linked immunosorbent assay tests, *Electrophoresis*, **2004**, 25(21-22): 3761-3768.
[45] Meisen, I., Peter-Katalinic, J., Muthing, J., Discrimination of neolacto-series gangliosides with alpha 2-3- and alpha 2-6-linked N-acetylneuraminic acid by nanoelectrospray ionization low-energy collision-induced dissociation tandem quadrupole TOF MS, *Anal. Chem.*, **2003**, 75(21): 5719-5725.
[46] Bindila, L., Almeida, R., Sterling, A., Allen, M., Peter-Katalinic, J., Zamfir, A., Off-line capillary electrophoresis/fully automated nanoelectrospray chip quadrupole time-of-flight mass spectrometry and tandem mass spectrometry for glycoconjugate analysis, *J. Mass Spectrom.*, **2004**, 39(10): 1190-1201.
[47] Vukelic, Z., Metelmann, W., Muthing, J., Kos, M., Peter-Katalinic, J., Anencephaly: Structural characterization of gangliosides in defined brain regions, *Biol. Chem.*, **2001**, 382(2): 259-274.
[48] Zamfir, A., Vukelic, Z., Peter-Katalinic, J., A capillary electrophoresis and off-line capillary electrophoresis/electrospray ionization-quadrupole time of flight-tandem mass spectrometry approach for ganglioside analysis, *Electrophoresis*, **2002**, 23(17): 2894-2903.
[49] Zamfir, A., Vukelic, Z., Bindila, L., Peter-Katalinic, J., Almeida, R., Sterling, A., Allen, M., Fully-automated chip-based nanoelectrospray tandem mass spectrometry of gangliosides from human cerebellum, *J. Am. Soc. Mass Spectrom.*, **2004**, 15(11): 1649-1657.
[50] Zamfir, A., Seidler, D. G., Kresse, H., Peter-Katalinic, J., Structural investigation of chondroitin/dermatan sulfate oligosaccharides from human skin fibroblast decorin, *Glycobiology*, **2003**, 13(11): 733-742.
[51] Zamfir, A., Vakhrushev, S., Sterling, A., Niebel, H. J., Allen, M., Peter-Katalinic, J., Fully automated chip-based mass spectrometry for complex carbohydrate system analysis, *Anal. Chem.*, **2004**, 76(7): 2046-2054.
[52] Williger, K., *doctoral thesis*, University of Bonn, 1993.

[53] De Laeter, J. R., Bohlke, J. K., De Bievre, P., Hidaka, H., Peiser, H. S., Rosman, K. J. R., Taylor, P. D. P., Atomic weights of the elements: Review 2000 - (IUPAC technical report), *Pure Appl. Chem.*, **2003**, 75(6): 683-800.

[54] Mamer, O. A., Lesimple, A., Letter to the editor, *J. Am. Soc. Mass Spectrom.*, **2004**, 15(4): 626-626.

[55] Ledford, E. B., Rempel, D. L., Gross, M. L., Space-charge effects in Fourier-Transform Mass-Spectrometry - mass calibration, *Anal. Chem.*, **1984**, 56(14): 2744-2748.

[56] Han, S. J., Shin, S. K., Space-charge effects on Fourier transform ion cyclotron resonance signals: Experimental observations and three-dimensional trajectory simulations, *J. Am. Soc. Mass Spectrom.*, **1997**, 8(4): 319-326.

[57] Taylor, P. K., Amster, I. J., Space charge effects on mass accuracy for multiply charged ions in ESI-FTICR, *Int. J. Mass Spectrom.*, **2003**, 222(1-3): 351-361.

[58] Zamfir, A., Lion, N., Vukelic, Z., Bindila, L., Rossier, J. S., Girault, H. H., Peter-Katalinic, J., Thin chip microsprayer system coupled to quadrupole time-of-flight mass spectrometer for glycoconjugate analysis, *Lab Chip*, **2004**, 5(3): 298-307.

[59] Bindila, L., Froesch, M., Lion, N., Vukelic, Z., Rossier, J. S., Girault, H. H., Peter-Katalinic, J., Zamfir, A., A thin chip microsprayer system coupled to Fourier transform ion cyclotron resonance mass spectrometry for glycoconjugate screening, *Rapid Comm. Mas Spectrom.*, **2004**, 18(23): 2913-2920.

[60] Froesch, M., Bindila, L. M., Baykut, G., Allen, M., Peter-Katalinic, J., Zamfir, A. D., Coupling of fully automated chip electrospray to Fourier transform ion cyclotron resonance mass spectrometry for high-performance glycoscreening and sequencing, *Rapid Commun. Mass Spectrom.*, **2004**, 18(24): 3084-3092.

[61] Metelmann, W., Peter-Katalinic, J., Muthing, J., Gangliosides from human granulocytes: A nano-ESI QTOF mass spectrometry fucosylation study of low abundance species in complex mixtures, *J. Am. Soc. Mass Spectrom.*, **2001**, 12(8): 964-973.

[62] Juraschek, R., Dulcks, T., Karas, M., Nanoelectrospray - More than just a minimized-flow electrospray ionization source, *J. Am. Soc. Mass Spectrom.*, **1999**, 10(4): 300-308.

[63] Lion, N., Gellon, J. O., Girault, H. H., Flow rate characterization of microfabricated microspray emitters, *Rapid Comm. Mass Spectrom.*, **2004**, 18(14): 1614-1620.

[64] McGuinness, P., Drenckhan, W., Weaire, D., The optimal tap: Three-dimensional nozzle design, *J. Phys. D: Appl. Phys.*, **2005**, 38(18): 3382-3386.

[65] Chen, H. H., Brenner, M. P., The optimal faucet, *Phys. Rev. Lett.*, **2004**, 92(16): 166106-1;166106-4.

[66] Heck, A. J. R., van den Heuvel, R. H. H., Investigation of intact protein complexes by mass spectrometry, *Mass Spectrom. Rev.*, **2004**, 23(5): 368-389.

[67] Vekey, K., Internal energy effects in mass spectrometry, *J. Mass Spectrom.*, **1996**, 31(5): 445-463.

[68] Drahos, L., Vekey, K., MassKinetics: a theoretical model of mass spectra incorporating physical processes, reaction kinetics and mathematical descriptions, *J. Mass Spectrom.*, **2001**, 36(3): 237-263.

[69] Gabelica, V., De Pauw, E., Karas, M., Influence of the capillary temperature and the source pressure on the internal energy distribution of electrosprayed ions, *Int. J. Mass Spectrom.*, **2004**, 231(2-3): 189-195.

Chapter 5. On-chip solid-phase extraction

1. Effects of salts on electrospray ionisation

It has been known since the earliest days of electrospray ionisation that the presence of salts is detrimental in terms of sensitivity and mass resolving power. Tang and Kebarle were among the first to study this issue [1]; they established an empirical law which shows that the electrospray current depends on electrolyte concentration in the bulk solution as:

$$I_{ES} = A\left(\lambda_m^0 [E]\right)^n \quad (1)$$

where λ_m^0 is the limiting molar conductivity, [E] the cation concentration and n a constant close to 0.22 (independent on the electrolyte nature). Furthermore, they observed that the analyte A^+ intensity in the mass spectrum varies with the electrolyte concentration as:

$$I(A^+) = Pf \frac{k_A [A^+]}{k_A [A^+] + k_E [E^+]} I_{ES} \quad (2)$$

where P is a sample factor related to the instrument used, f is the fraction of ions converted into the gas phase, $[A^+]$ and $[E^+]$ are analyte and cation supporting electrolyte concentrations, k_A and k_E are rate constants expressing the rate of transfer of analyte and electrolyte ions to the gas phase. Unfortunately, equation (2) did not fit measured analyte intensities over a wide concentration range: a good fit was obtained in the low concentration range of $[A^+]$ ($<10^{-5}$ M) with the assumption that $k_A/k_E=1$, but deviated from experimental results above 10^{-5} M.

In further studies [2, 3], Tang and Kebarle showed that at low concentration k_A/k_E relates directly to relative surface activities of A^+ and E^+, whereas at high analyte concentrations, k_A/k_E expresses the relative surface activities *and* relative ion evaporation rates, provided that the ion evaporation model of Iribarne holds true [4]. However, they were unable to introduce a unified model that accounts for both concentration regions.

1.1. The partition model

More recently, Enke introduced a model for matrix and analyte effects in electrospray ionisation, based on the hypothesis that ion intensities in a mass spectrum directly relates to ion abundances on electrosprayed droplets surfaces [5]. Basically, an electrospray droplet is formed with an excess charge Q that is the net difference between anions and cations. In positive ion mode (Q>0), cations partition between the bulk droplet, in which they ion-pair with anions, and the surface of the droplet. One clear limitation of Enke's approach is that this partition is postulated *a priori*, and is only validated by consistency with experimental results (see below). It is thus important to investigate how ions distribute in charge droplets.

Gouy-Chapman in electrospray droplets

Following the approach of Myland and Oldham [6], let consider an electrospray droplet containing cations of charge number z^+ and anions of charge number z^-, bearing a total charge Q and having a radius R. The electric field inside a droplet at distance r from its centre (taken as a potential origin) is given by Gauss's law:

$$-\left(\frac{d\phi}{dr}\right)(r) = \frac{Q(r)}{4\pi\varepsilon r^2} \qquad for\ 0 \le r \le R \qquad (3)$$

where ϕ is the local electrical potential within the droplet, Q(r) is the charge within a sphere of radius r, and ε is the solution's permittivity. Let $n^+(r)$ and $n^-(r)$ be the cation and

anion number densities respectively (in m^{-3}), z^+ and z^- be the cation and anion number of charge, and q(r) the volumetric charge density (in $C.m^{-3}$). Then comes:

$$e\left[z^+n^+(r)+z^-n^-(r)\right]=q(r) \tag{4}$$

where *e* is the elementary charge ($1.6\ 10^{-19}$C). According to Boltzmann's statistics,

$$\begin{aligned} n^+(r) &= n_0^+ \exp\left(-z^+\frac{e}{k_BT}\phi(r)\right) \\ n^-(r) &= n_0^- \exp\left(-z^-\frac{e}{k_BT}\phi(r)\right) \end{aligned} \tag{5}$$

where n_0^+ and n_0^- are cation and anion number densities at r = 0, and Q_e/k_BT the Boltzmann factor (38.9215 V^{-1} at 298K). Contrary to the classical Gouy-Chapman derivation, where the potential origin is fixed in the bulk neutral solution, far from the charged planar surface, in order to preserve the generality of the derivation for very small droplets whose radius is not large compared to the Debye length, it is not postulated here that the global charge density is zero at the centre of the droplet. Poisson's equation in spherical coordinate writes:

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\phi}{dr}\right)=\frac{-q(r)}{\varepsilon} \tag{6}$$

Incorporating equation (5) into (6) gives:

$$\frac{d}{dr}\left(r^2\frac{d\phi}{dr}\right)=\frac{-er^2}{\varepsilon}\left[z^+n_0^+\exp\left(-z^+\frac{e}{k_BT}\phi(r)\right)+z^-n_0^-\exp\left(-z^-\frac{e}{k_BT}\phi(r)\right)\right] \tag{7}$$

This is the Poisson-Boltzmann equation to be solved with the following boundary conditions:

$$\left(\frac{d\phi}{dr}\right)_{r=R}=\frac{-Q}{4\pi\varepsilon R^2} \qquad \text{(Gauss's law at r = R)} \tag{8}$$

$$\left(\frac{d\phi}{dr}\right)_{r=0}=0 \qquad \text{(no potential discontinuity at the centre of the droplet)} \tag{9}$$

Myland and Oldham solved equation (7) numerically, by a finite difference model [6] for a droplet at the limit of Rayleigh fission, meaning that its radius R and global charge Q are related by:

$$Q = 8\pi\sqrt{\varepsilon_0 \gamma R^3} \tag{10}$$

Where γ is the surface tension of the liquid used. What they found is the cation and anion distributions within the droplet, as shown in Figure 66:

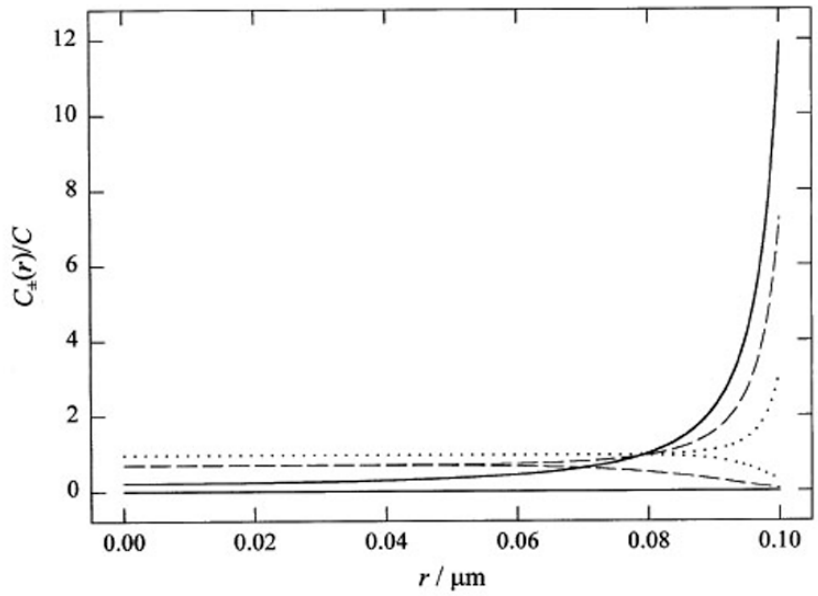


Figure 66. For a droplet of radius 0.1 μm, the curves show concentration profiles normalised by the feed concentration of the 1:1 electrolyte solution. The full, dashed and dotted lines relate to concentrations of 0.02482, 0.06572 and 0.21036 eq.m^{-3}. In all cases, the upper curve represents the cation and the lower curve the anion. Reprinted from [6] with permission.

As clearly visible in Figure 66, most cations are close to the droplet surface; it can be estimated from the graph and other results presented in [6] that 75% of cations are contained in the 10% outer shell, whereas 95% of anions are contained in the 90% inner

shell. This outer cation amount is even bigger when (i) the electrolyte concentration increases and (ii) when the droplet size increases.

Enke's simple partition model

The calculation of electrolyte distributions shown above now gives a basis for the two-compartment model adopted by Enke [5], in which cations partition between a bulk phase in the core of the electrospray droplet where they are neutralised by anions, and a surface layer, as shown below:

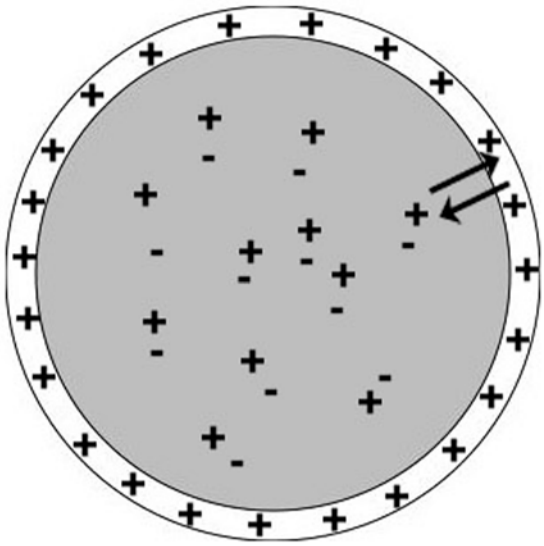


Figure 67. Scheme of the two-compartment model used in the partition model

This model is a macroscopic model (that do not describe physical reality) comparable to the ones often used in biosensor data analysis, where diffusion processes toward the surface of the biosensor are fitted by a two-compartment model, such as in [7, 8]. In the case studied here, cations can be analyte ions A^+ or supporting electrolyte E^+, whereas anions are counterions X^-. At this point, the thickness of the outer shell need not be defined. Partition of analyte and electrolyte ions between the two compartments can thus be described by:

$$\left(A^+X^-\right)_i \rightleftarrows \left(A^+\right)_S + \left(X^-\right)_i \qquad K_A = \frac{\left[A^+\right]_s\left[X^-\right]_i}{\left[A^+X^-\right]_i} \tag{11}$$

$$\left(E^+X^-\right)_i \rightleftarrows \left(E^+\right)_S + \left(X^-\right)_i \qquad K_E = \frac{\left[E^+\right]_s\left[X^-\right]_i}{\left[E^+X^-\right]_i} \tag{12}$$

Where s subscripts denote surface compartment and i subscripts the interior compartment of the droplet. The excess charge Q born by the droplet is then :

$$[Q] = \left[A^+\right]_s + \left[E^+\right]_s \tag{13}$$

Given that $[Q]=I_{ES}/F\Gamma$ (14) where F is the Faraday constant and Γ the flow rate, in typical electrospray conditions with I_{ES} = 100 nA and Γ = 6 µl/min, the excess charge is approximately 10^{-5} M. If we assume that the total fraction of analyte and electrolyte present in the outer compartment is very small (which is contradictory to paragraph "Gouy-Chapman in electrospray droplets", see above), then K_A and K_E reduce to:

$$K_A = \frac{\left[A^+\right]_s\left[X^-\right]_i}{C_A} \qquad K_E = \frac{\left[E^+\right]_s\left[X^-\right]_i}{C_E} \tag{15}$$

Where C_A and C_E are the analyte and electrolyte bulk concentrations. If $[X^-]_i$ is eliminated from equations (15):

$$\left[A^+\right]_s = \frac{C_A K_A}{C_E K_E}\left[E^+\right]_s \tag{16}$$

Which together with equation (13) results in:

$$\left[A^+\right]_s = \frac{C_A K_A}{C_A K_A + C_E K_E}[Q] \tag{17}$$

Equation (17) relates the analyte concentration in the outer compartment (which has been hypothesised to define the mass spectrum intensity) to the excess charge concentration present in the electrospray droplets. Remarkably, this relation is very similar to equation (2)

introduced by Tang and Kebarle [1], even though it was derived from very different principles; Tang and Kebarle used ion evaporation rates whereas the model presented above uses only equilibrium constants. However, equation (17) suffers from the same limitation as Tang and Kebarle's model, which was unable to account for nonlinearity when analyte intensity in the mass spectrum was plotted versus analyte concentration in the bulk solution. We shall thus consider a more general model, where no *a priori* hypothesis is made about analyte and electrolyte ratios between the two compartments.

Enke's general partition model

The mass conservation of analyte and electrolyte ions can be written as:

$$\begin{aligned} C_A &= \left[A^+\right]_s + \left[A^+X^-\right]_i \\ C_E &= \left[E^+\right]_s + \left[E^+X^-\right]_i \end{aligned} \tag{18}$$

Then from equation (11) and (12) comes:

$$\frac{K_A}{K_E} = \frac{\left[A^+\right]_s \left(C_E - \left[E^+\right]_s\right)}{\left(C_A - \left[A^+\right]_s\right)\left[E^+\right]_s} = \frac{\left[A^+\right]_s \left(C_E - Q - \left[A^+\right]_s\right)}{\left(C_A - \left[A^+\right]_s\right)\left(Q - \left[A^+\right]_s\right)} \tag{19}$$

Equation (19) can now be rearranged to reach a quadratic equation in $[A^+]_s$:

$$\left[A^+\right]_s^2 \left(\frac{K_A}{K_E} - 1\right) - \left[A^+\right]_s \left([Q]\left(\frac{K_A}{K_E} - 1\right) + C_A \frac{K_A}{K_E} + C_E\right) + C_A [Q] \frac{K_A}{K_E} = 0 \tag{20}$$

This equation relates directly $[A^+]_s$ and C_A with use of only the partition equilibrium constants K_A and K_E, and the measurable excess charge [Q]. At high values of analyte concentration C_A, equation (20) simplifies into $[A^+]_s$ = [Q], which accounts for the saturation region experimentally observed. At very low values of C_A, equation (20) simplifies into:

$$\left[A^+\right]_s = C_A \left(\frac{K_A / K_E}{K_A / K_E - 1 + C_E / [Q]}\right) \tag{21}$$

Which accounts for the linear region observed experimentally at low analyte concentration.

From equation (20), one can now plot the analyte concentration in the outer compartment versus the analyte concentration in the bulk electrosprayed solution, as shown in Figure 68:

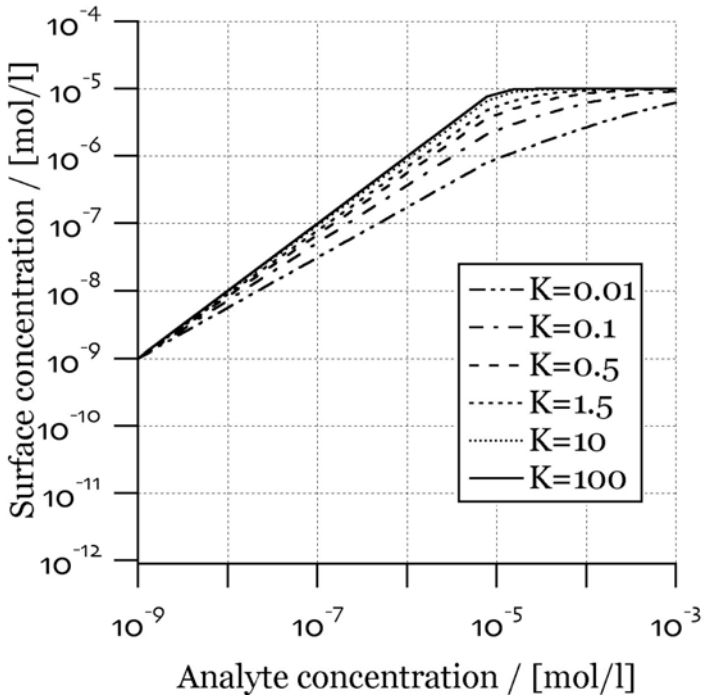


Figure 68. Response function for a singly charged analyte with a singly charged electrolyte. [Q] = 10^{-5} *equiv/l,* $C_E = 10^{-3}$ *M and* $K=K_A/K_E=0.01$ *to 100.*

In can be seen from Figure 68 that the mass spectrometric response for a given analyte is highly dependant on the K_A/K_E ratio, which means that analyte and electrolyte ions compete to reach the outer compartment (surface of the droplet): at high K_A/K_E ratio (analyte with a much stronger surface affinity that the electrolyte), the analyte intensity is perfectly linear up to the saturation regime, whereas at low K_A/K_E ratio (electrolyte with a much stronger surface affinity that the analyte), the mass spectrometric response is non-linear. This model was validated by Enke [5] on data from Kebarle and Tang [9], as shown in Figure 69:

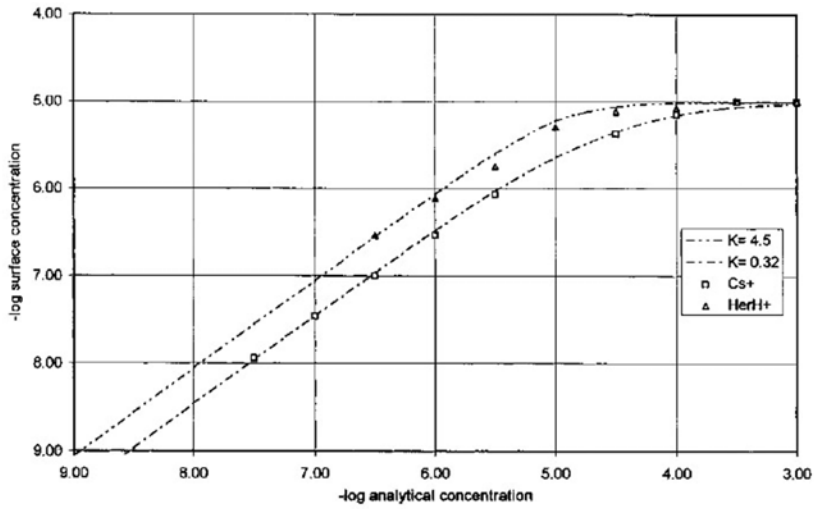


Figure 69. Fit of equation (20) on data presented by Kebarle and Tang in [9]: electrospray analysis of heroin hydrochloride ($HerH^+$) and cesium (Cs^+) ions in the presence of 10^{-5} M of NH_4CH_3O. Reprinted with permission from [5].

Effect of salt concentration in Enke's model

The model presented above allows to predict the response for analyte A^+ in the presence of various concentration of salts, based on equation (20):

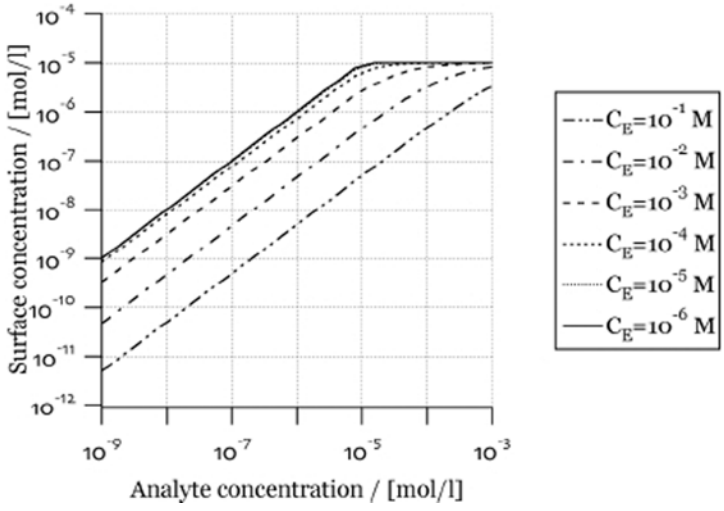


Figure 70. Calculated response for a singly-charged analyte at various salt concentrations. ($K_A/K_E=50$, $[Q]=10^{-5}$ equiv/l)

The presence of salts is clearly responsible for a loss of sensitivity: for an analyte concentration of 10^{-7} M, switching from 10^{-6} to 10^{-1} M of salts results in a decreased analyte intensity by more than two orders of magnitude. However, comparison of this model with experimental results revealed a more complicated situation [10]:

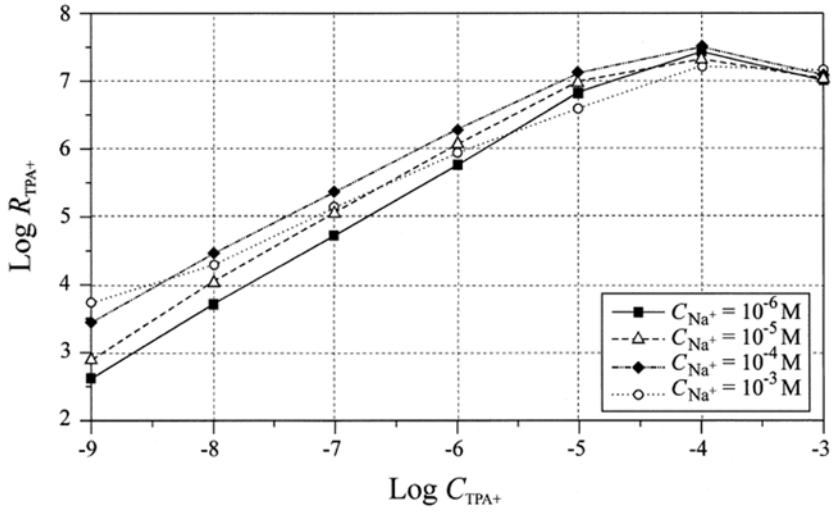


Figure 71. Effect of tetrapentylammonium (TPA^+) and sodium ions concentrations on the intensity of TPA^+ in the mass spectrum. Reprinted from [10] with permission.

In agreement with the model, when $C_{Na}{}^+ < 10^{-3}$ M, the TPA^+ response is proportional to $C_{TPA}{}^+$ up to 10^{-5} M. However, for a given TPA^+ concentration, the response increases with increasing $C_{Na}{}^+$ up to 10^{-4} M and then decreases again, which is in contradiction with the model. In fact, Enke's partition model assumes that [Q] is independent of analyte and electrolyte concentrations, which is not true; in particular, [Q] increases with increasing electrolyte concentration:

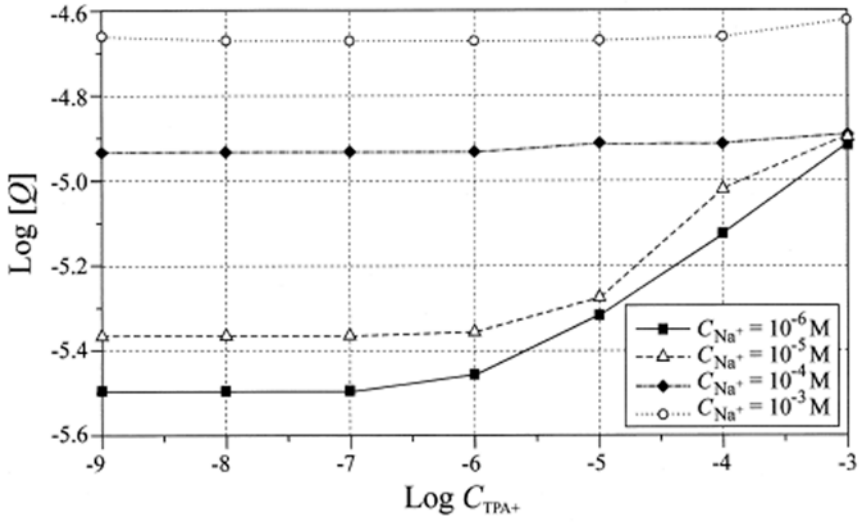


Figure 72. Effect of C_{TPA}^{+} and C_{Na}^{+} on [Q], which was calculated from eq. (14) using the measured I_{ES} and Γ = 8µl/min. Reprinted with permission from [10].

Unfortunately, taking into account this dependency of [Q] on electrolyte concentration does not explain the effect of increasing electrolyte concentration, and to our knowledge, no clear explanation on the analyte intensity dependency on electrolyte concentration was provided to date [10].

But it is worth noting that the partition model accounts at least qualitatively for the increased sensitivity, dynamic range and better robustness of nanospray ionisation towards the presence of salts [11]: the total excess charge [Q] is inversely proportional to the flow rate Γ (see equation (14)), which means that in classical electrospray conditions where Γ=6 µL/min and I_{ES}=100 nA, [Q]≈10^{-5} eq/l whereas in nanospray conditions where Γ=50 nL/min and unchanged current, [Q]= 10^{-3} eq/l. The available excess charge is thus much higher in nanospray compared to classical electrospray, as shown in Figure 73:

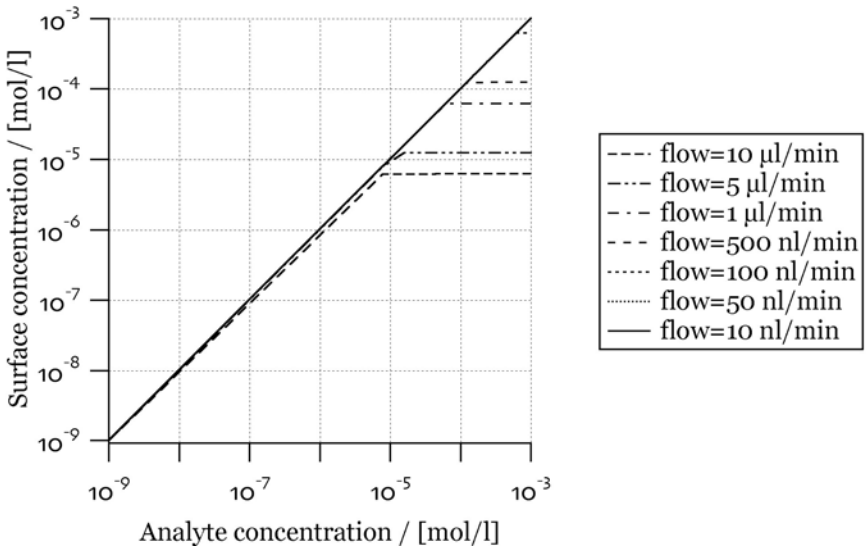


Figure 73. Response curves at different flow rates for an electrolyte concentration of 10^{-5} M, K_A/K_E=50, I_{ES}= 100 nA.

More interestingly, at low flow rates, the effect of increasing salt concentrations is less pronounced, as shown in Figure 74:

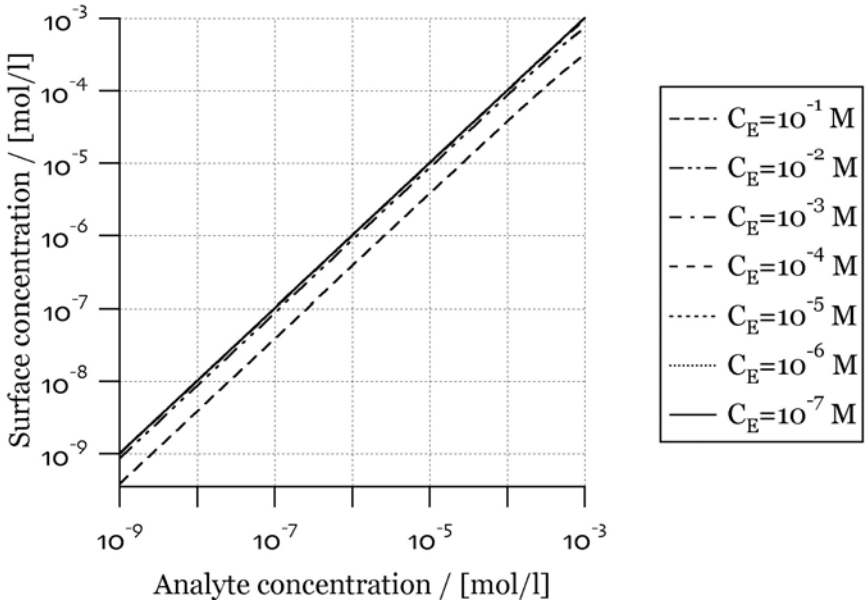


Figure 74. Effect of salt concentration on analyte response for a flow rate of 50 nl/min, K_A/K_E=50, I_{ES}= 100 nA.

When comparing Figure 74 to Figure 70 for which Γ=6 µl/min, one can see that the decrease in analyte response when increasing the electrolyte concentration from 10^{-6} to 10^{-1}

M passes from more than two orders of magnitude (for classical electrospray) to less than half an order of magnitude (for nanospray).

Pros and cons of Enke's partition model

Enke's partition model was successful in describing the saturation observed when increasing analyte concentration (see for example Figure 69); this saturation of analyte intensity is not due to instrumental limitations, but to intrinsic properties of electrospray ionisation: the global charge available is indeed limited by the electrospray current itself (see equation (14)). The model also qualitatively accounts for the decrease of sensitivity observed with increasing salt concentrations. To date, it has been the most successful investigation and prediction of mass spectrometric response and has been extended qualitatively to the competition of several analytes for ionisation [5, 12]. However, the whole model is based on two strong hypotheses:

1. the intensity of a compound in the mass spectrum is proportional to its concentration at the surface of electrospray droplets
2. compounds partition between the bulk of electrospray droplets and an outer compartment.

which physical meaning is not so clear. Hypothesis 1 implies that gas phase ions are formed solely by ion evaporation from electrospray droplets [2, 13-17], which is thought to hold true for small compounds [18]. For analytes larger than 3.3 kDa, the charge residue model [15, 19, 20], which hypothesises that electrospray droplets undergo coulombic fissions and solvent evaporation till only one analyte molecule be present in the droplet, prevails. In particular, it better accounts for protein charge states observed in ESI-MS [21-24]. All experimental validations of Enke's partition model were conducted with small inorganic (see Figure 69) or organic ions (see Figure 71). It is highly doubtful that this

model accounts for mass spectrometric response of large analytes such as polypeptides and proteins.

Moreover, it is broadly accepted that charged analytes locate at the surface of electrospray droplets because of electrostatic repulsions, which is at the core of hypothesis 1. However Znamenskiy *et al* showed by very simple calculations that metastable cases can occur, in which a potential energy minimum corresponds to charges located at the centre of the droplet [25]. They even found by molecular dynamics calculation that ions are not located close to the surface, as shown in Figure 75:

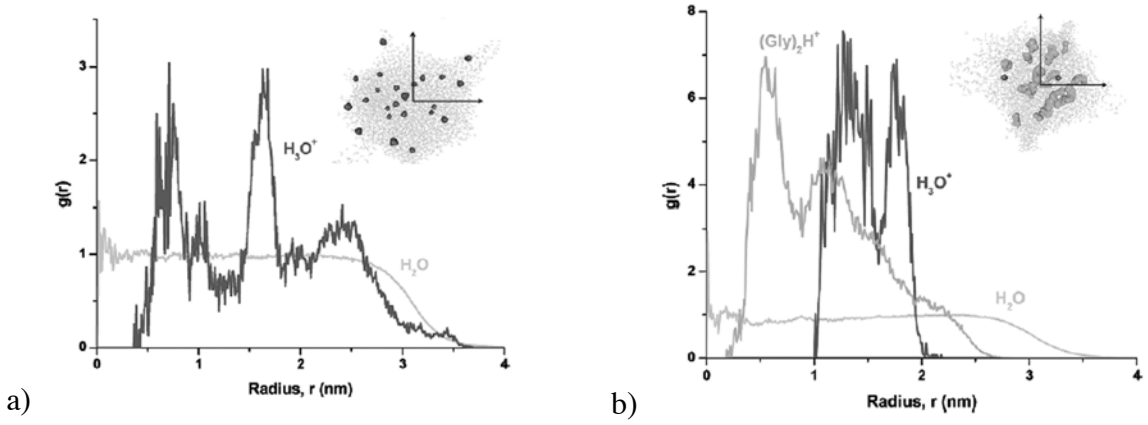


Figure 75. Radial distribution of water molecules and ions around the centre of the nanodroplet. a) the droplet contains 23 hydronium ions b) the droplets contain 21 diglycine ions and 2 hydronium ions. Reprinted with permission from [25].

Hypothesis 2 states that ions partition between the core of the droplet and an outer shell. Whereas this may not be true, as shown in the Gouy-Chapman derivation of ion distribution in charged droplets (see paragraph "Gouy-Chapman in electrospray droplets" above) the physical meaning of the partition coefficient is not clear. Its first aim is to reflect differences in compounds surface affinities, which comprises both electrostatic and solvation effects.

Moreover, the partition model does not take into account droplet evolution through time (solvent evaporation, droplet fission, ion evaporation). Recently, Cech *et al* tried to

refine the partition model by incorporating a simple fissioning process [26]: when solvent evaporates, the droplet shrinks till its surface reaches 85% of its initial volume and then fissions. Offspring droplets are then thought to carry 2% of the parent droplet's mass and 15 % of its charge, as assumed by Gomez *et al* [27]. The model is thus re-evaluated taking into account the number of fissions undergone by electrospray droplets. Unfortunately, it does not account for experimental results any better than the partition model presented above and in [5, 10].

Finally, none of the above-mentioned approaches take into account the external electric field, which has a typical value of 10^6 $V.m^{-1}$. For the sake of comparison, the electric field just outside a droplet carrying 100 elementary charges is plotted in Figure 76a: the "intrinsic" field is generated by the droplet charge itself (calculated by Gauss' law) and the external field is the one corresponding to the electrospray voltage. Comparably, the intrinsic field for droplets at their Rayleigh limit (see equation (10)) is plotted in Figure 76b.

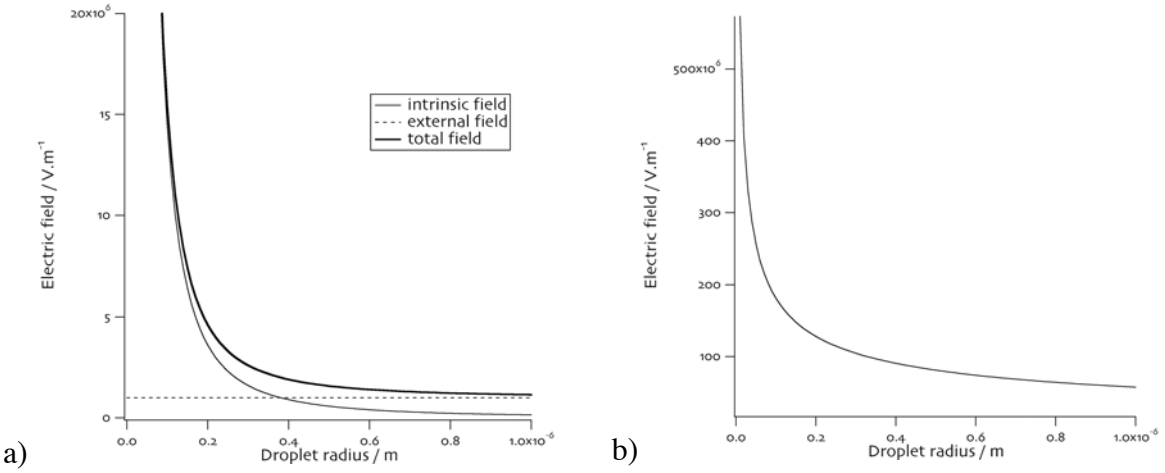


Figure 76. Values of the electric field just outside a) droplets carrying 100 elementary charges, and b) droplets at their Rayleigh limit.

As can be seen from Figure 76, if the number of charges is fixed arbitrarily to 100, the external electrical field has a non-negligible value for bigger droplets (>200-400 nm in radius), whereas if one considers droplets at their Rayleigh limit, the intrinsic field is always

much higher than the external one. It thus appears reasonable to neglect the external electric field in the study of fissioning droplets.

1.2. Analyte cationisation by salts

The previous section deals with the effect of salts on global response of electrospray ionisation. However, the presence of salts in the analysed solution can result in another effect with direct consequences for mass spectrometric analyses: analyte cationisation. Briefly speaking, analyte ionisation during electrospray is driven first by an in-solution charging, defined by the analyte pK_a:

$$A + H_3O^+ \leftrightarrows AH^+ + H_2O \qquad (22)$$

In a second stage, a gas-phase analyte charging can occur, defined by the analyte proton affinity:

$$A + H^+ \leftrightarrows AH^+ \qquad (23)$$

The proton affinity (PA) is defined as the negative change of enthalpy associated with the protonation reaction (23).

But gas-phase charging can also occur via “metallization”, for example by sodium ions:

$$A + Na^+ \leftrightarrows ANa^+ \qquad (24)$$

Which is thus defined by sodium affinity. The direct consequence of metallization is the appearance of extra peaks in the mass spectrum: besides the $[M+nH]^{n+}$ distribution, $[M+nNa]^{n+}$ and possibly $[M+nH+mNa]^{(n+m)+}$ and $[M+nNa-mH]^{(n-m)+}$ distributions appear, which is detrimental to the mass spectrometric analyses: it can first decrease sensitivity by dividing the analyte intensity into several peaks, and complicate mass calculation and spectrum deconvolution. Figure 77 shows the complexity that can arise from the presence of salt clusters in the spectrum of a simple cyclic pentapeptide:

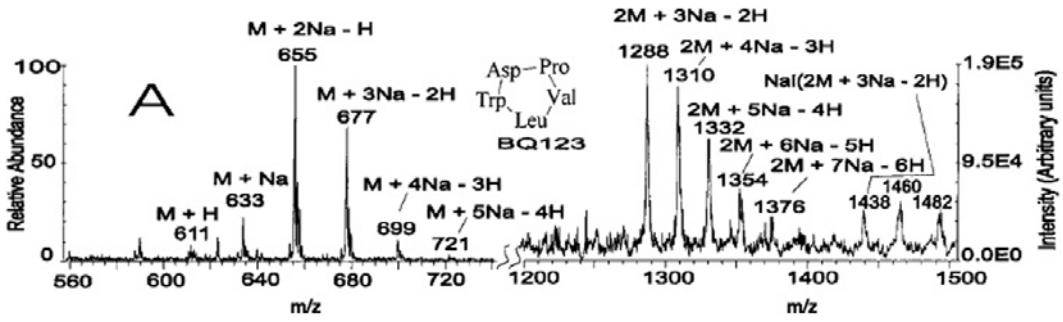


Figure 77. Mass spectrum of a single cyclic pentapeptide in the presence of sodium salts. Reprinted with permission from [28].

Fortunately, peptides and proteins usually have much higher affinity for protons than for sodium cations. For example, Feng *et al* used the Cooks kinetic method [29] to measure lithium and sodium ion binding energies of amino acids [30]; in this methodology, complexes of the form $P_1M^+P_2$ are formed by electrospray ionisation, and further dissociated in the instrument, typically by collision activated dissociation. The reactions undergone by the peptide complexes are thus the following:

$$P_1M^+P_2 \xrightarrow{k_1} P_1M^+ + P_2$$

$$P_1M^+P_2 \xrightarrow{k_2} P_2M^+ + P_1$$

$$P_2M^+ + P_1 \leftrightarrows P_1M^+ + P_2 \qquad K$$

$$P_1M^+ \rightarrow P_1 + M^+ \qquad \Delta G_M(1)$$

$$P_2M^+ \rightarrow P_2 + M^+ \qquad \Delta G_M(2)$$

The observed intensities can thus be converted into thermodynamic values by:

$$\frac{I\left(P_1M^+\right)}{I\left(P_2M^+\right)} = \frac{k_1}{k_2} \approx K$$
$$-RT_{eff} \ln K = \Delta G_M(1) - \Delta G_M(2) \quad (25)$$

This technique thus allows to calculate differences in binding energies between pairs of peptides. Using this technique as well as reference experiments, Feng *et al* found that proton affinities of dipeptides were in the range of 10^6 J/mol whereas their sodium affinities were in the range of 10^5 J/mol. This factor of 10 in binding energies corresponds to a factor of 22000 in equilibrium constants, and thus to a much stronger affinity of peptides for protons than for sodium.

2. State-of-the-art in microfluidic sample clean-up

From the discussion above, it can be clearly understood that sample clean-up prior to electrospray ionisation is of utmost importance. Sample clean-up comprises desalting, but also elimination of co-solvents usually encountered in biological sample preparation protocols, such as solubilisation agents (detergents, urea, thiourea…). Moreover, as electrospray ionisation most often requires the presence of organic solvents for efficient analyte desolvation, sample clean-up devices that allow solvent exchange as well are highly desirable. Microfluidic sample clean-up have been reviewed in references [31-33].

2.1. Sample clean-up by solid-phase extraction

Solid-phase extraction (SPE) refers to processes where analytes of interest are retained selectively on a stationary phase, while solvents and undesired compounds are washed out, and retained analytes can therefore be eluted [34]. SPE is particularly promising in combination with microfluidic devices, because elution volumes can be

incomparably small. Typical features of solid-phase materials are their retention mechanism (hydrophobicity, electrostatic interactions, affinity…) and their specific surface area.

With wall derivatisation

One straightforward approach to incorporate a solid-phase in a microfluidic system is to derivatise microchannel walls with molecular reagents that bear the desired functional property. The biggest advantage of this approach is the ease of functionalisation, as it is usually accomplished by flowing reagents through the microchannel to be derivatised. However, its biggest limitation is that the surface-to-volume ratio of the solid-phase is defined by microchannel geometry only. Kutter *et al* have presented a glass microchip whose walls have been derivatised with octadecyltrimethoxysilane to make a C18 solid phase [35]. Whereas extraction performances were demonstrated with a fluorescent dye alone in simple buffers, the only limitation of this approach is the limited specific surface area that can be reached (from 230 to 781 mm^{-1}, as reported in [36]), compared to other solid-phase systems (see Table 14 at the end of this section). However for some applications the high surface-to-volume ratio of a microfluidic system may be sufficient to achieve a proper SPE with wall derivatisation only.

Based on reverse-phase microparticles

One of the conceptually easiest ways to integrate a solid-phase into a microfluidic system is to use microbeads; unfortunately, packing a microchannel with microbeads turns out to be a difficult technical task. First the packing material must be flown and trapped into the microchannel; Li *et al* [37] used a first plug of 40 µm C18 beads, to make a filter at the junction of a large 800 µm wide, 150 µm deep and 22 mm long channel and a small 10 µm deep, 30 µm wide injection channel. The large channel was then packed with 5 µm C18 beads, resulting in a surface-to-volume ratio around 107 m^{-1} (estimated from authors' data

and Waters' technical information). This system reached nanomolar sensitivity for leucine-enkephalin, when coupled to on-chip CE and direct infusion to ESI-MS. The same team previously reported an elegant methodology to pack microchannels with beads: restriction microstructures were integrated in the microchannel as shown in the left panel of Figure 78:

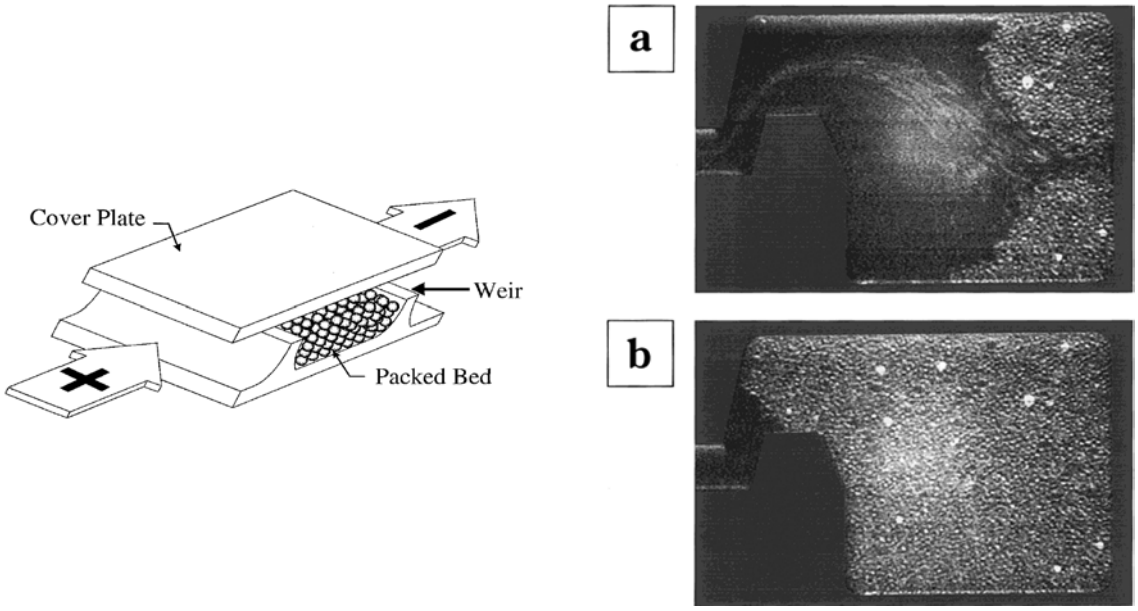


Figure 78. A 330 pL microchamber is defined by two weirs in the sample flow path that retain microparticles but allow solution flow. Left: schematic diagram of the device; the transverse microparticle filling channel is not shown. Reprinted with permission from [38] Right: a) image of the chamber at an initial stage of electrokinetic packing and b) after completion. Reprinted with permission from [39].

Two weirs were microstructured in the flowing path of the sample channel, so that the gap between the top of the weirs and the cover plate that closes the microchannel is 1 μm; this two weirs define a 330 pL microchamber that can be filled with C18 silica particles (from 1.5 to 4 μm in diameter), under electroosmotic flow using a transverse channel (not shown in left panel of Figure 78) [38]. Panel a) shows an early stage of packing, and panel b) the packing after completion. Moreover, the authors reported a so-called "solvent lock": electrokinetic packing is performed in acetonitrile, and when aqueous buffer is applied to the reverse-phase bed, particles tend to "lock". With this set-up, authors were able to

preconcentrate and detect 1 pM solutions of fluorescent molecules after a 3 min loading of the reverse-phase bed [39]. More recently, Marko-Varga *et al* [40, 41] used a simplified "weir approach" to pack a 330 µm wide and 200 µm deep microchannel with a 100-300 nL reverse-phase bed. This microfluidic device was used to remove salts and urea from standard peptide mixtures or from in-gel digested proteins prior to deposition on MALDI targets. 1 nM peptide mixtures (20 fmol) were successfully analyzed as well as various proteins identified after 2D-GE and in-gel digestion. Despite these few successful uses of microparticles, packing a microchannel with reverse-phase microbeads remains a difficult task, because of clogging, back pressure effects, interactions of particles with microchannel walls… Lettieri *et al* have recently introduced a new concept for bead handling in microfluidic systems [42]: when opposing a parabolic pressure driven flow profile with a flat EOF profile (with a low net flow rate) at converging or diverging channels, recirculation of microbeads occurs, so that trapping can be performed without any mechanical restriction, as shown in Figure 79.

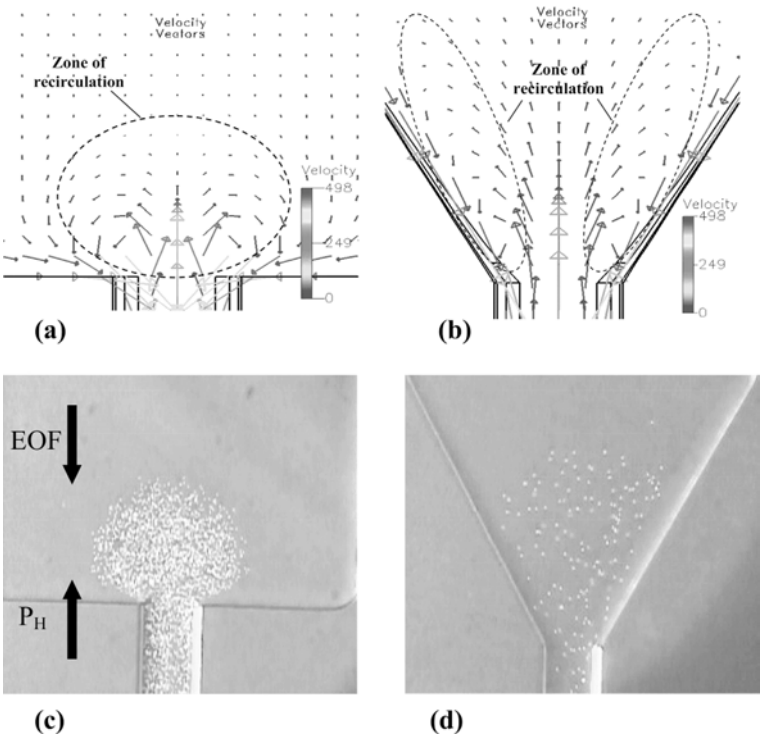


Figure 79. Protein A coated beads (1 µm) circulation at a divergent channel under opposite EOF and pressure-driven flow. (a), (b) Numerical simulations of vector flow (µm/s^{-1}) with two different geometries; (b), (c) corresponding experiments. Reprinted from [42] with permission.

Furthermore, if one of the flows is shut down, microbeads can then be moved to another part of the microfluidic device (e.g. a detection window). This methodology was assessed with the capture of fluorescent biotin by streptavidin beads, and washing out of excess biotin. This work shows a nice example of an innovative solution to trap microparticles in a microfluidic device. It is thought that many other methodologies will appear in the future for microparticle handling in microchannels.

Based on monoliths

As discussed above, the two classical ways to integrate solid-phases in microchannels are (i) wall derivatisation, which leads to limited surface-to-volume ratios and makes compulsory the use of small channels or (ii) bead packing, which usually leads

to back-pressure and clogging problems. An elegant and innovative solid-phase integration has been pioneered by Frechet's group: monolithic materials are polymerised directly in the microchannel, after the channel has been filled with the monomers mixture. Polymerisation is then induced whether by temperature or by UV light. The choice of monomers allows the design of the desired surface chemistry (polarity, hydrophobicity, charge, reactive groups for grafting antibodies or proteins) and the experimental condition (mainly the composition of the porogenic solvent) allows the control of pore size. For example Yu *et al* have tested the use of photopolymerisable monoliths based on methacrylate derivatives [43, 44]: ethylene methacrylate (EDMA), 2-hydroxyethyl methacrylate (HEMA), butyl methacrylate (BMA), 3-(trimethoxysilyl)propyl methacrylate, glycidyl methacrylate (GMA), were used to produce monoliths with different surface properties. Whether ion exchange or hydrophobic solid-phases were incorporated into glass microchips, with a mean pore size of 19 µm to allow flow-through experiments, and tested as SPE sorbents with a model peptide and Green Fluorescent Protein. Both analytes underwent a preconcentration of 1000 when adsorbed on a 7-mm-long bed (100 µm wide, 40 µm deep) with a surface-to-volume ratio of 106 to 107 m^{-1} (estimated from authors' data) and eluted with a variety of salts or acetonitrile/water mixtures respectively. These simple proofs of principle exemplify the ease of integration of porous monoliths in microfluidic devices as well as the versatility of surface properties that can be tailored.

Solid phase type	**Specific surface-to-volume ratio**	**From ref.**
Wall derivatisation	781 mm^{-1}	[36]
Microparticles	107 m^{-1}	[38, 39]
Monoliths	107 m^{-1}	[43, 44]

Table 14. Comparison of surface-to-volume ratios obtained with different solid-phase preparation strategies.

2.2. Filtration and microdialysis

Dialysis systems have been incorporated into microfluidic devices by several teams, most of time for direct coupling with ESI-MS. The simplest device consists in two microchannels placed face-to-face and sandwiching a dialysis polymeric membrane. Sample is then flown through the membrane to exclude high-molecular weight compounds as in [45], or counter flown against a dialysis buffer, so that salts and low molecular weight compounds are expelled [46], as shown in Figure 80.

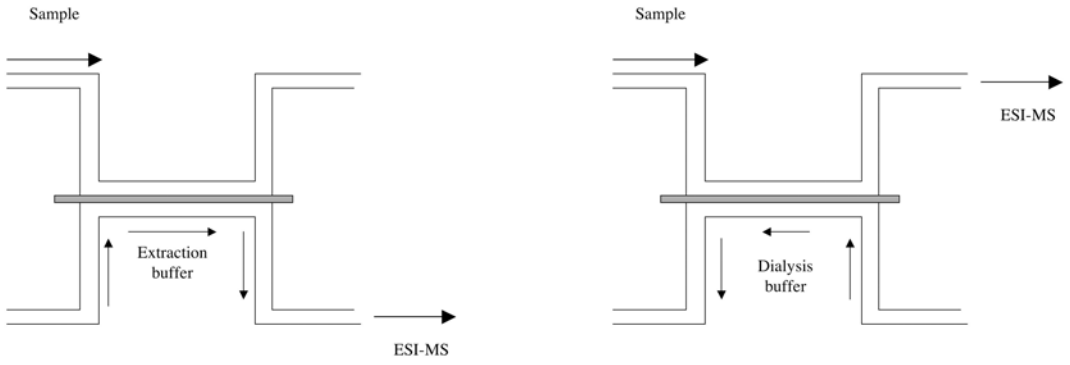


Figure 80. The two schemes introduced for sample filtration (left) or microdialysis (right)

Xiang *et al* [47] developed a dual dialysis miniaturised system: the device embeds one high- and one low molecular-weight-cut-off membrane, and the sample is sequentially counter flown against both of them before being injected into ESI-MS, which allowed the analysis of a whole *E. Coli* lysate prepared in 6.8 M urea. Jiang *et al* [48] proposed a somewhat more refined dialysis and filtration microdevice: in a first stage, a mixture of aflatoxin antibody and aflatoxin is flown against a dialysis buffer at a PVDF membrane interface, so that all unbound aflatoxin is expelled; the complexes are then flown against dry air to be concentrated, and enter a second dialysis device, where they are dissociated; the eluted aflatoxins can then be analyzed by ESI-MS. The use of microfluidic dialysis allowed a gain in sensitivity of two orders of magnitude in the detection of aflatoxins, thanks to

limited dilution during dialysis, post-dialysis concentration and limited loss of sample through integration of all steps.

2.3. Liquid-liquid extraction

Liquid-liquid extraction is a well-established procedure for sample preparation in bioanalytics [49], due to its robustness and specificity; in the recent years, it has regained interest due to its potential for high-throughput [50]. Many protocols have been developed, whether based on aqueous two-phase systems (often salt / detergents systems) or aqueous / organic solvent multiphase systems. This can be applied to enrichment or isolation of subclasses of proteins or isolation of single proteins directly from biological matrices [51].

At the same time, microfluidic systems are operated under laminar flow (low Reynolds number), which means that two solutions can be flown next to each other within the same channel without any other mixing than analyte diffusion from one flow to another [52]. Up to now, multiphase extraction microdevices have been applied only in inorganic chemistry: Tokeshi *et al* have performed extraction of complexed cobalt in an aqueous/ toluene dual-flow microchannel. A very stable 2 mm long interfacial zone was established and the flow was stopped. The extraction was then completed within 60s whereas the conventional solvent extraction procedure (starting from 10 ml aqueous solution) was ca. 10 min long [53]. Schilling *et al* presented a more integrated approach for cell lysis, protein extraction and β-galactosidase quantitation [54]: cells-containing solution and lysis buffer are flown under pressure driven flow in the same channel, so that cell lysis occurs at the flows interface. Cell debris essentially do not diffuse in the lysis buffer, whereas protein components partition between the two flows, as shown in Figure 81.

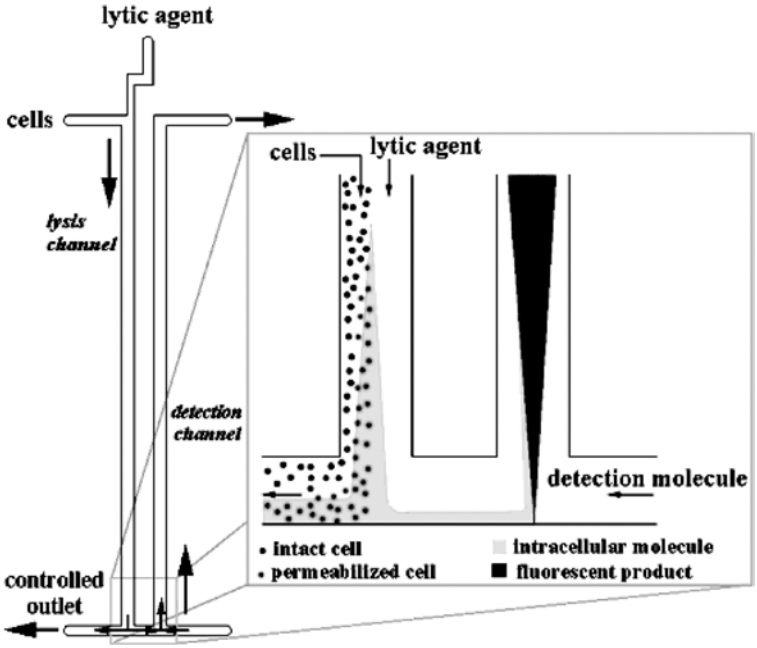


Figure 81. Design of the microfluidic device for cell lysis, protein extraction and β-galactosidase detection. Reprinted with permission from [54].

The lysis buffer is then flown in another microchannel along with a detection flow containing a fluorogenic substrate for β-galactosidase; fluorescence appears at the flows interface, so that the enzyme can be quantified. This rather simple device allows the isolation of intracellular compounds from cell debris, and it is likely that fractionation properties (range of extracted molecules' diffusion coefficients) can be adjusted by changing the flow rates: by increasing flow rates, large molecules will have less time to diffuse in the lysis flow, so that only low-molecular weight compounds will be extracted.

3. On-chip solid-phase extraction as stand-alone device

As shown above, solid-phase extraction is the most favoured technique for sample clean-up prior to mass spectrometric analysis, both for microfluidic devices [35, 37-41, 43, 44] and in classical sample preparation protocols, as exemplified by the success of functionalised pipette tips such as ZipTips™ from Millipore. However, integrating a solid-

phase into microchannels turns out to be a difficult technical task: first if a packing approach is to be used, one needs to integrate a frit into microchannels to retain microparticles, which is feasible only with photolithographic techniques; then the assembled microfluidic structure has to be pressure-resistant, so as to cope with back pressures usually encountered in packing protocols. On the other hand, developing solid-phase monoliths imposes compatibility of the polymerization solvents with the microchip material, and grafting the solid-phase onto microchannels walls to avoid any shrinking effect during polymerization.

3.1. Polymeric membranes as a solid-phase

On the contrary, one can use polymeric membranes for off-line capture of analytes of interest, and then interface them at the inlet of a microfabricated polymer microspray, as shown below:

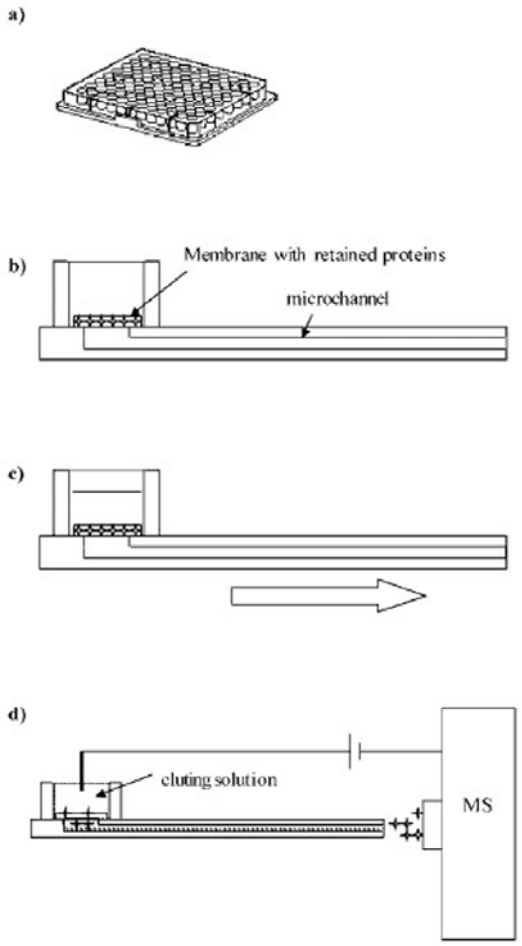


Figure 82. Scheme of operation: membranes are first incubated in the sample to clean-up (a), then placed at the bottom of the microspray reservoir (b); the membrane is then washed by pumping deionised water through the microspray (c), and the elution solution is then applied, together with the electrospray voltage (d).

3.2. Material and methods

This experimental scheme was tested with poly(vinylidene difluoride) (PVDF) membranes Immobilon P (0.45 µm pore size, 120 µm thick) from Millipore which are usually employed in electroblotting experiments of proteins after 1D or 2D gel electrophoresis [55, 56]; they retain compounds primarily through hydrophobic interactions, and to some extent through electrostatic effects [57]. Model compounds were S-propranolol hydrochloride from Fluka (Buchs, Switzerland), insulin from bovine pancreas and equine myoglobin from Sigma (Saint-Louis, MO, USA), and human angiotensin II from Bachem (Bubendorf, Switzerland). Phosphate buffer saline (PBS) tablets (0.138 M NaCl, 0.0027 M

KCl, 0.01 M phosphate, pH 7.4) were from Sigma, HPLC grade methanol from Merck (Darmstaadt, Germany), and acetic acid from Fluka.

Microsprays were produced by DiagnoSwiss SA (Monthey, Switzerland) in polyimide, with microchannel dimensions of 120 µm wide, 45 µm deep and 1 cm long, and closed in-house with a polyethylene / polyethylene terephthalate (PE/PET) 35 µm sheet (Morane, Oxon, UK) using a laminator (Morane, Oxon, UK) set at 130 °C (upper roll), 50 °C (lower roll) and 2 bars. They were cut in a tip shape with scissors, and a reservoir was pasted over the inlet.

All experiments were done on a LCQ Duo mass spectrometer (Thermo Electron, San Jose, CA, USA) used in positive ion mode polarity; the electrospray source was removed, and the microchip positioned in front of the heated capillary inlet as previously reported [58, 59]. The electrospray voltage was delivered by a platinum electrode plunged into the microchip reservoir, and was around 2.4 kV. All acquisitions were done in full scan mode ($50 \leq m/z \leq 2000$) or MS/MS mode without any averaging. The heated capillary was kept at 200 °C, and in each experiment, ion optics parameters were optimised on the compound of interest.

Membranes are first cut in 6 mm diameter pieces, which fit the microchip reservoir as well as the bottom of flat 96 well microtiter plate wells, and wet with methanol and rinsed with milliQ water, as recommended by Millipore. They are then incubated in PBS with 10% methanol (v/v) with the molecule of interest, under shaking at room temperature, whether for one hour or overnight. The incubation volume is typically 200 µl in a 96 well plate. In a second step, the membrane is placed within the reservoir over the microchannel inlet, and rinsed extensively with milliQ water by pumping 250 µl three times through the microchannel. The membrane is then wet with the spraying solution (MeOH:acetic acid:

water at 50:1:49%) and placed in front of the MS heated capillary. The high voltage is then applied by placing a platinum electrode within the reservoir.

3.3. Results

Detection of propranolol from a salty solution

The present desalting scheme has been successfully applied to the detection of a small drug, propranolol, in PBS + 10% methanol, without any previous treatment of the sample. The membrane was incubated one hour in a 500 µg/ml propranolol solution, at room temperature under shaking, and consecutively washed and wet by 50 µl of the spraying solution (50% methanol : 49% water : 1% acetic acid). Figure 83 shows the spectrum of propanolol eluting from the membrane, with the single peak at m/z = 260.14 and no contaminant coming from the polymer microchip or the membrane.

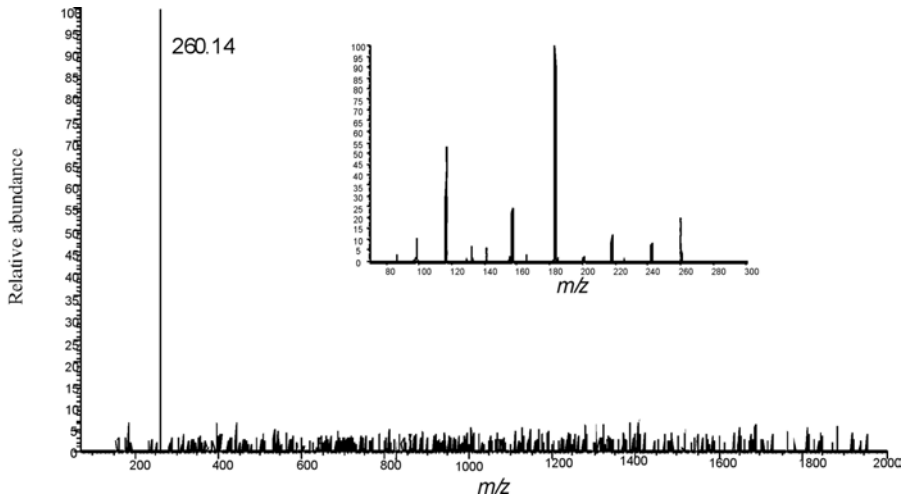


Figure 83. MS spectrum of propranolol eluting from the membrane incubated in 500 µg/ml (1.7 mM) in PBS. Inset: MS/MS spectrum of parent ion m/z 260.14 at 30% collision energy.

The inset in Figure 83 shows the MS/MS spectrum of parent ion m/z = 260.14, whose peaks are listed in Table 1. MS/MS was acquired at 30% collisional energy.

m/z	Relative intensity
98.14	9.63
116.24	49.32
132.14	1.86
157.15	23.53
183.16	100
218.02	10.24
242.11	10.29
260.14	22.15

Table 15. m/z list of ions in the MS/MS spectrum of parent ion m/z = 260.14 of the propranolol MS spectrum

In comparison, Figure 84 shows the spectrum of a 500 µg/ml propranolol solution sprayed directly from PBS.

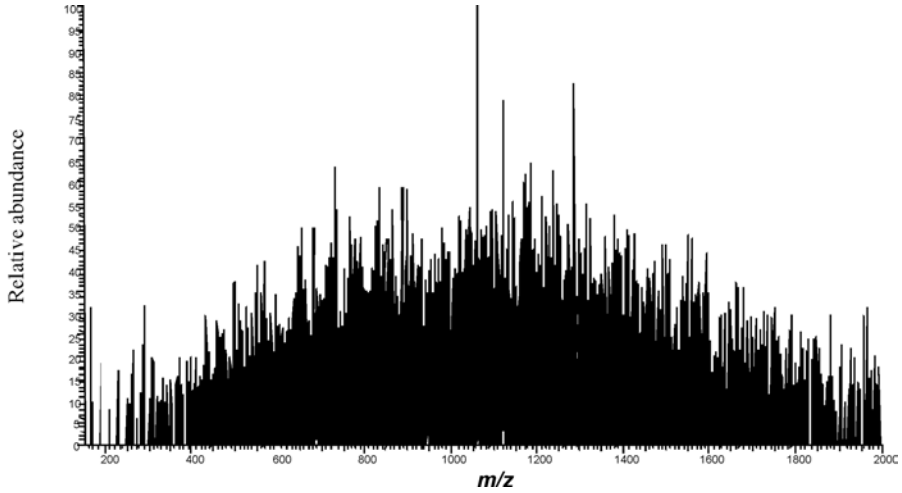


Figure 84. MS spectrum obtained from a direct electrospray of 500 µg/ml propanolol in PBS + 50% methanol (vol:vol).

In order to investigate the effect of the concentration of propranolol in the incubation solution, membranes were incubated in various concentrations of propranolol, from 500 µg/ml down to 10 ng/ml (1.7 mM down to 33.8 nM). In order to avoid any misleading identification of propranolol, the quantitation was made on the main peak of the MS/MS spectrum of parent ion *m/z* = 260.14, that is to say *m/z* = 183.16, which is shown in Figure 85. There is a clear relationship between the logarithmic concentration and the

logarithmic intensity of the peak. But even 500 µg/ml is far from the saturation of the membrane binding capacity for propranolol in this system.

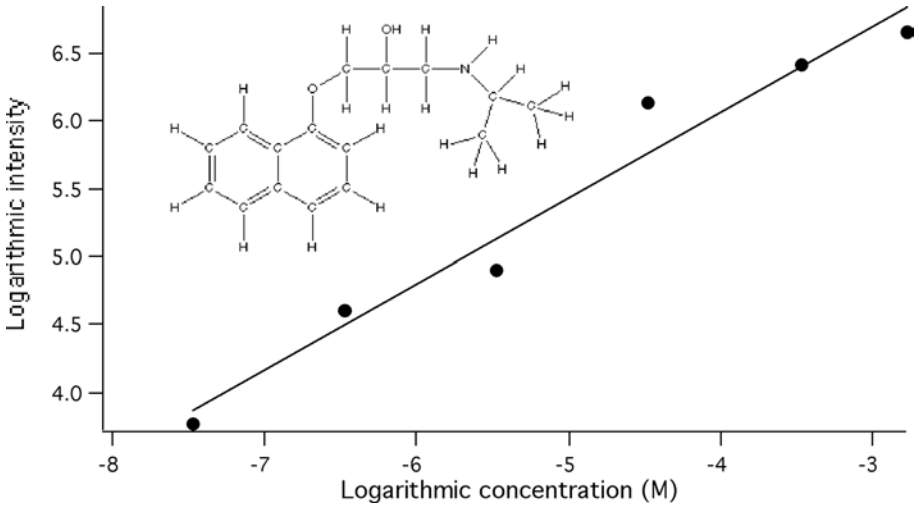


Figure 85. Intensity of the m/z=183.1 main peak of the MS/MS spectrum of propanolol (parent ion m/z=260.1) at various incubation concentrations (from 10 ng/ml to 500 µg/ml).

Peptide and protein desalting

The procedure described above has also been applied successfully to human angiotensin II. In particular, the electrospray from a microchannel outlet has been reported to be very stable through time without the membrane [58]. Electrospray generation through the membrane does not impend the temporal stability: Figure 4 shows the reconstructed chromatogram of specific mass *m*/*z*=523.8 (main peak of human angiotensin II) in the solution eluting from a membrane incubated for one hour in 10 µg/ml (9 µM) solution of angiotensin. The 50 µl of spraying solution added over the membrane in the chip reservoir could be sprayed during more than half an hour without any problem, which shows that the presence of the membrane does not impend the physical stability of the spray. Of course most of the adsorbed angiotensin is released from the membrane within the first five minutes, due to disruption of hydrophobic interactions between the membrane and angiotensin molecules by the spraying solution. In principle, the elution duration could be

adjusted if needed by increasing or decreasing the percentage of methanol in the spraying solution.

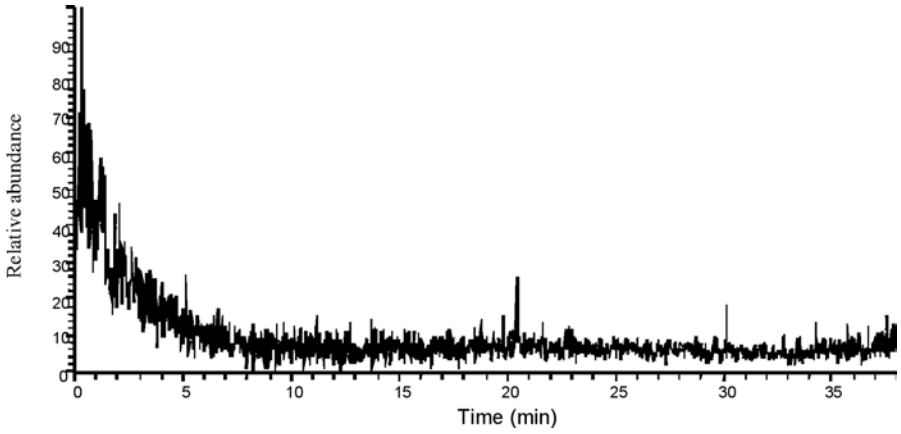


Figure 86. Ion reconstructed "chromatogram" of ion m/z 523.8.

Insulin was also desalted: membranes were incubated for one hour at room temperature in 200 µl of different concentrations of insulin (from 0.2 to 75 µM) in PBS +10% methanol. Figure 87 (left) shows the spectrum of insulin eluting from a membrane incubated in a 10 µM insulin saline solution. The quantitation of insulin eluting from the membranes was done on the $(M+5H)^{5+}$ ion (m/z = 1147.5): Figure 87 (right) shows the logarithmic absolute intensity of the peak versus the logarithmic incubation concentration. In order to estimate the saturation concentration, an exponential was fitted to the experimental points. Saturation was estimated as 95% of the plateau value, which gives a saturation concentration of 40 µM in this system.

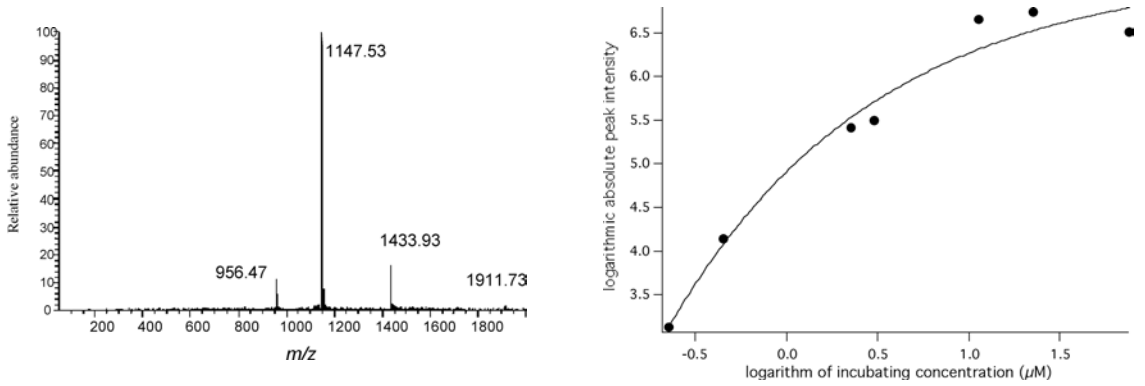


Figure 87. Desalting of insulin. Left: spectrum of insulin eluting from the membrane. Right: Intensity of peak m/z = 1147.5 versus concentration of insulin in the starting sample.

Furthermore, in order to validate the procedure with larger molecules, membranes were incubated in a 10 µM myoglobin solution, for one hour, at room temperature, in 200 µl of PBS + 10% methanol. The membrane was consecutively washed, put in a microspray, and 50 µl of the spraying solution was put over the membrane. The resulting spectrum is shown in Figure 7; no salt cluster can be observed nor contaminant coming from the membrane.

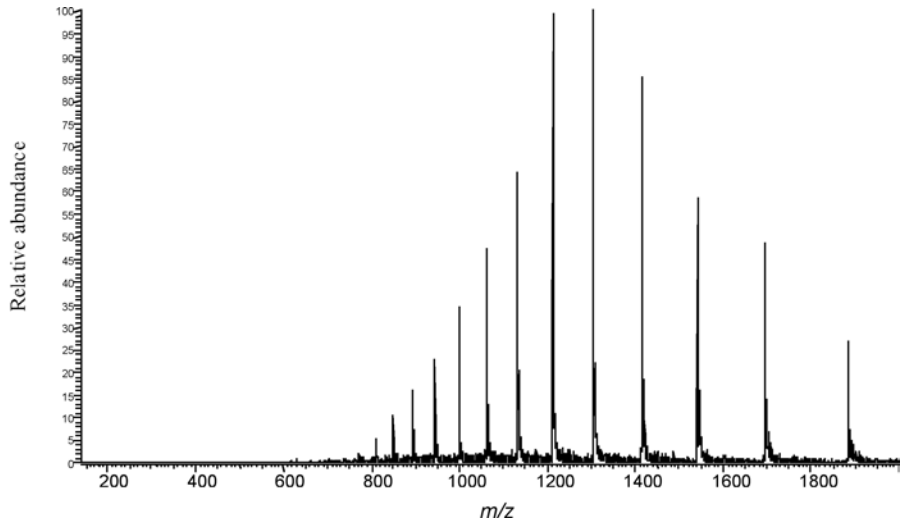


Figure 88. Desalting of myoglobin from a PBS solution. Single spectrum of solution eluting from the membrane.

Influence of elution solution composition

In order to gain some insight into the dependency of the solvent composition on elution behaviour, membranes were incubated in 10 μM of insulin in PBS, placed on the chip, and washed with milliQ water; different elution solutions were then applied, and the total ion chromatograms were recorded:

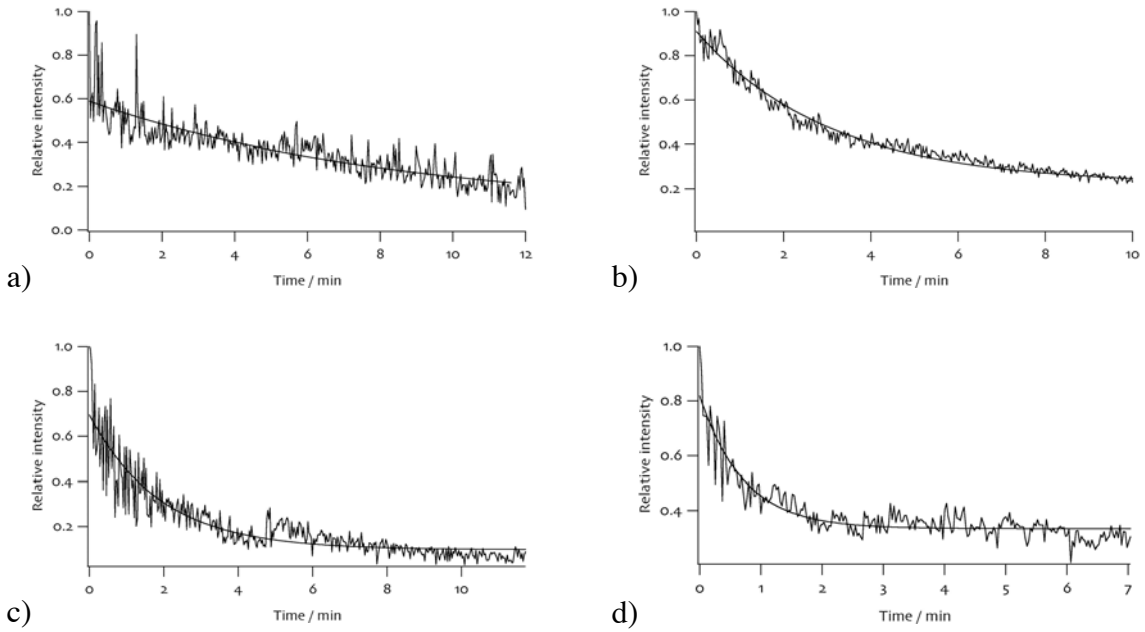


Figure 89. Total ion current of insulin eluting from the membrane in a) 30% MeOH: 1% acetic acid: water, b) 40% MeOH: 1% acetic acid: water, c) 60% MeOH: 1% acetic acid: water, d) 70% MeOH: 1% acetic acid: water

In order to measure the elution rate, total ion chromatograms were fitted to:

$$TIC = TIC_0 + A\exp\left(-\frac{t-t_0}{\tau}\right) \qquad (26)$$

Which gave the elution rate τ for each solvent composition:

% of methanol	Rate constant τ (min^{-1})
30	8.51±1.60
40	3.06±0.07
60	1.91±0.10
70	0.71±0.04

Table 16. Effect of methanol content on elution behaviour

4. On-chip solid-phase extraction with infusion

In order to minimise the elution duration by controlling externally the elution solution flow rate, a capillary connection was added to the solid-phase extraction set-up, as shown below:

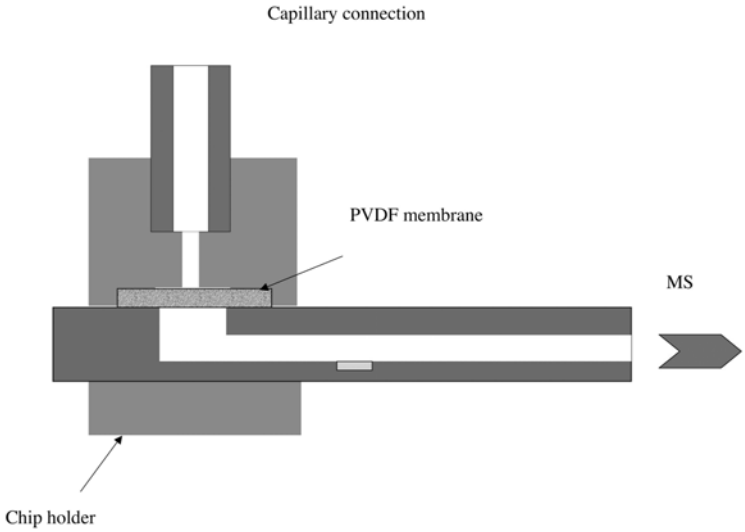


Figure 90. Scheme of on-chip SPE with capillary connection.

4.1. Material and methods

Urea, S-propranolol hydrochloride and cytochrome C from horse heart were obtained from Fluka (Buchs, Switzerland). Phosphate buffered saline (PBS) tablets (0.138 M NaCl, 0.0027 M KCl, 0.01 M phosphate, pH 7.4), 1,4-dithio D,L-threitol (DTT), iodoacetamide, CHAPS (3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate), β-lactoglobulin A from bovine milk and insulin from bovine pancreas were from Sigma (St Louis, USA). Human angiotensin II was from Bachem (Bubendorf, Switzerland). Methanol (Merck, Darmstadt, Germany) and acetic acid (Fluka) were used without any further purification. Deionised water (18.2 MW) was produced using a Milli-Q system from Millipore (Bedford, USA). Proteins were dissolved in deionized water to a 1 mg/ml

concentration. The PVDF membrane is an Immobilon-P™ membrane from Millipore (Bedford, USA).

Polyimide microchips were provided by DiagnoSwiss (Monthey, Switzerland). Their fabrication has been described in detail elsewhere [60, 61]. The microchannel is 120 µm wide, 35 µm deep and 1 cm long. The coupling of the microchip with the mass spectrometer is quite simple, and has been presented previously [58, 59, 62]: one extremity of the microchannel is cut in a tip shape, and placed in front of the MS heated capillary. The spraying voltage (typically between 2 and 2.8 kV) is applied through a microelectrode embedded within the microchannel, and the position of the chip is adjusted to obtain a stable spray. The microchip is coupled to a syringe pump (Cole Parmer 74900) thanks to a home made low dead-volume holder. The PVDF membrane is sandwiched between the inlet of the microchannel and the fluidic connection, as shown in Figure 90. The mass spectrometer is a LCQ Duo Ion Trap (Thermo Electron, San Jose, USA). The heated capillary is kept at 200 °C, and the spray voltage is adjusted in order to have a spray current between 50 and 200 nA. All spectra are acquired in the positive mode, in full-scan mode.

The membrane is first wet with methanol (0.2 µl/min during 5 min) and then rinsed with milliQ water (0.2 µl/min during 5 min), as recommended by the membrane supplier. The sample is then adsorbed on the membrane by pumping the salty solution through the membrane at 0.2 µl/min to 5 µl/min. When the desired volume is pumped through the membrane, salts are washed out by pumping water (same duration and flow rate compared to the retention step). The eluting and spraying solution (50% methanol: 1% acetic acid: water) is then pumped to elute and spray the retained drugs, peptides and proteins.

β-lactoglobulin A was incubated in PBS (pH 7.4) at a concentration of 2.7 µM in the presence of 8 M urea and 4 mM DTT for 30 minutes at 45 °C. Just before incubation, the

solution was gently flushed with nitrogen, in order to expel oxygen. Iodoacetamide was then added to a final concentration of 38 mM, as in a previous study [63]. The solution was flushed again with nitrogen, and the mixture was incubated at 45 °C in the dark for 30 minutes. DTT and iodoacetamide stock solutions were freshly prepared.

4.2. Results

Sample preconcentration

In order to investigate the preconcentration that can be obtained, propranolol dissolved in PBS to a concentration of 87 µM was pumped at 0.5 µl/min during 30 minutes. After washing out of salts with deionised water, the spraying solution was applied at 0.2 µl/min. The position of the chip was optimized and the flow rate was then switched to 0.6 µl/min. Figure 91 shows the elution profile of propranolol. Spectra were acquired in full-scan mode without any averaging, and the chromatogram of specific ion m/z = 260.3 was post processed.

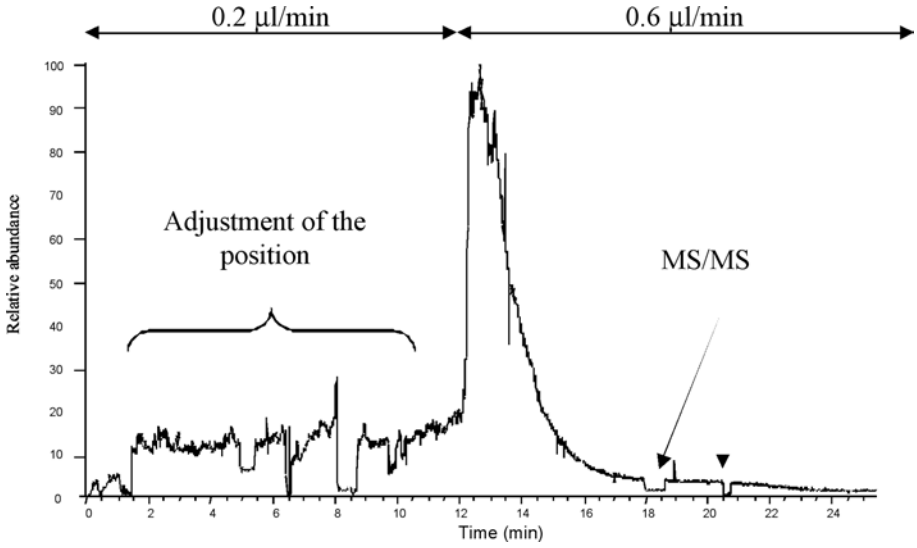


Figure 91. Reconstructed chromatogram of propranolol ion.

It can be observed that approximately 90% of propanolol is eluted within 2 minutes, which results in a preconcentration ratio of 12.5. The tailing that can be seen on the elution profile in Figure 91 can be attributed to diffusive phenomena within the membrane, and desorption of analytes into the reservoir, which represents a dead volume of 5 nL. It can be noted that the elution behaviour of the analyte depends on the eluant flow rate, on the nature of the eluant, and on the concentration of organic solvent within the eluant. All three parameters can be optimised for each analyte of interest. For example to preconcentrate a very hydrophobic analyte, one can use acetonitrile instead of methanol, optimise the concentration of acetonitrile in order to recover the maximum amount of analyte from the membrane, and then increase the eluant flow rate until the sampling rate of the mass spectrometer becomes the limiting factor.

Desalting of a mixture

A mixture of propranolol (50 µM), insulin (50 µg/ml) and cytochrome C (50 µg/ml) dissolved in PBS was pumped through the microchip at 0.2 µl/min during 15 minutes. Salts were then washed out with water, and the eluting and spraying solution was applied at 0.6 µl/min. Figure 92 shows the resulting spectrum: all analytes were recovered and no salt cluster can be observed. The noise level is comparable with that obtained for salt-free solutions.

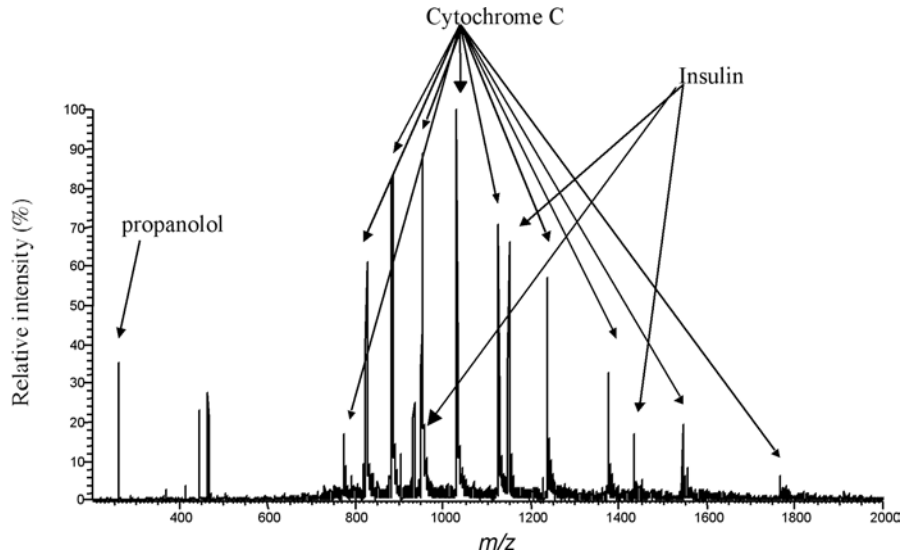


Figure 92. Desalting of a mixture of propranolol, insulin and cytochrome C dissolved in PBS.

Desalting in the presence of urea

Urea is used classically as a solubilising agent for hydrophobic proteins [64] or as a denaturant. A mixture of propanolol (50 µM), insulin (50 µg/ml) and cytochrome C (50 µg/ml) dissolved in PBS in the presence of 7M urea was pumped through the microchip at 0.2 µl/min. The elution step was done at 0.6 µl/min. Figure 93 shows the resulting spectrum, recorded without any averaging.

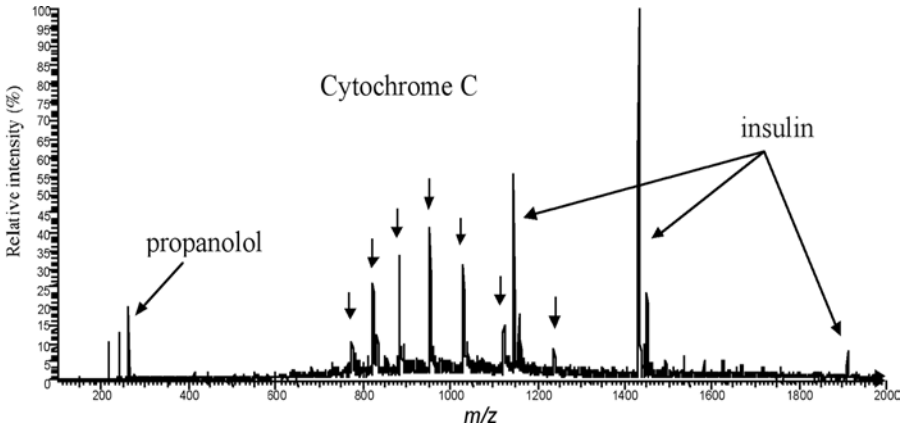


Figure 93. Desalting of a mixture of propranolol, insulin and cytochrome C dissolved in PBS in the presence of 7M urea.

Surprisingly, the presence of urea at high concentration does not prevent the adsorption of analytes on the PVDF membrane and does not add any spectral noise. Whereas the mechanism of action of urea as a protein denaturant is not fully understood, it is well established that urea disrupts intra-molecular hydrophobic interactions. But it does not seem that urea prevents interactions between hydrophobic analytes and the PVDF membrane.

Interestingly, one can see from Figure 92 and Figure 93 that the cytochrome C peak distribution is shifted towards lower *m*/*z* ratios when incubated in high concentrations of urea. It is now well established that the peak distribution in ESI mass spectra is strongly correlated with the folding state of the protein [65-73]: whereas it is not clear yet whether shifted charge states arise from a shielding of protonation sites in well-packed hydrophobic cores of folded proteins, or if in unfolded proteins, protonated sites can not be neutralized by structurally neighbouring oppositely charged sites [23], it is consensual that unfolded proteins have charge state distributions shifted to lower *m*/*z* ratios compared to native proteins. Here, one can see in Figure 93 that cytochrome C denatured in 7M urea does not refold during washing out of salts and elution, as shown by the loss of higher peaks attributed to the native state of the protein compared to Figure 92.

Desalting in the presence of detergents

Detergents are commonly used for hydrophobic proteins solubilisation [64, 74]; though the rationale for selectivity of particular detergents for sub-classes of proteins solubilisation is not clear, classical detergents (or detergent mixtures) are used in proteomics. The most commonly used detergents in protein biochemistry are shown in Figure 94:

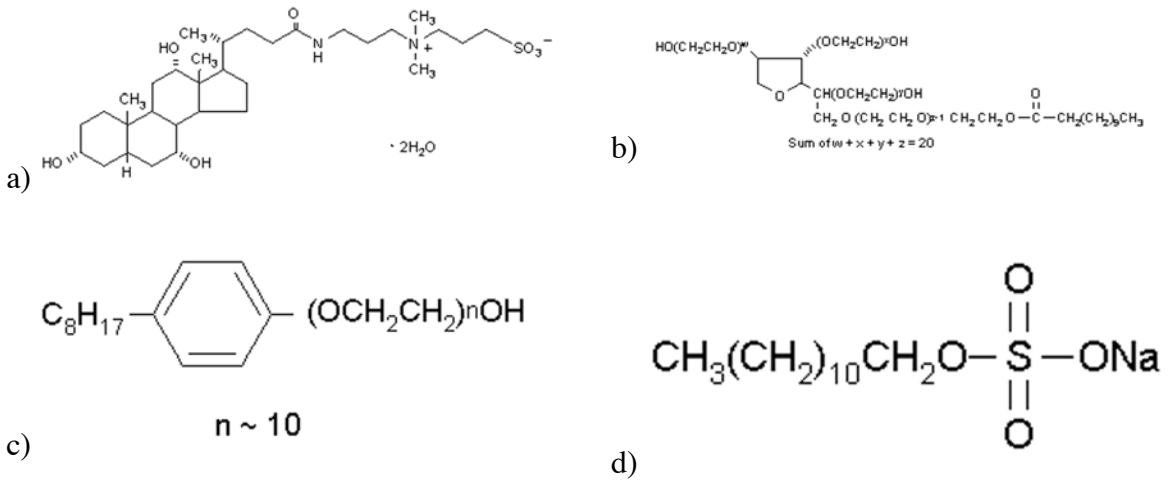


Figure 94. Commonly used detergents in protein biochemistry: a) CHAPS (zwitterionic) b) Tween 20 (non ionic) c) Triton X100 (non ionic) d) SDS (anionic)

The on-chip sample clean-up was tested in the presence of CHAPS; detergent concentrations of 0.1%, 1%, 2% and 4% (w:v) were added to a mixture of 50 µg/ml of human angiotensin II, 50 µM of propranolol, 50 µg/ml of insulin, and 50 µg/ml of cytochrome C in PBS. At all concentrations tested, it was impossible to recover any signal from propranolol, angiotensin and cytochrome C, as shown in Figure 95:

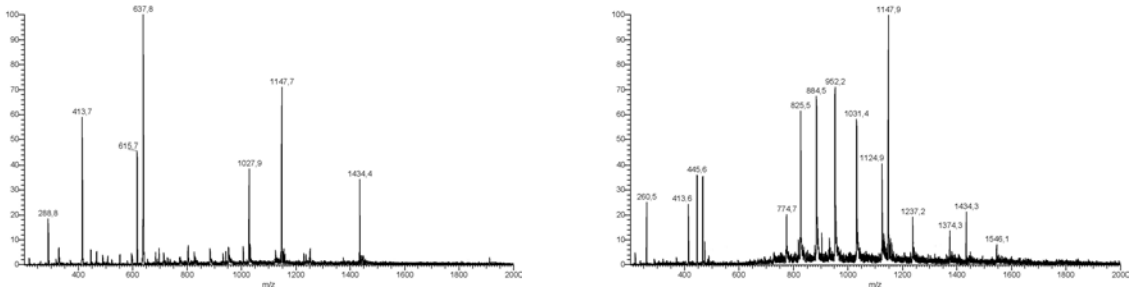


Figure 95. Sample clean up in the presence of CHAPS. Left: 0.1% of CHAPS (w:v) added to the sample. Right: control experiment with no CHAPS added.

Only insulin and the heme group from cytochrome C are visible in the left spectrum (peaks at *m/z* 288, 413, 637, and 1027 are contaminants present in the instrument); the rationale for this is that these are the most hydrophilic compounds present in the sample

Chemical tagging of cysteine residues

In order to show the interest of such an integrated desalting step, a specific chemical tagging of cysteine residues of β-lactoglobulin A was done with iodoacetamide. The scheme of the reaction is shown below:

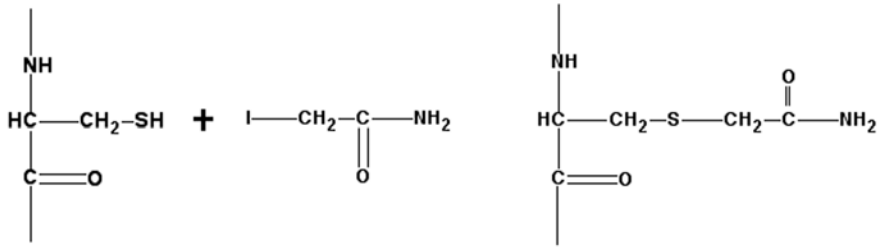


Figure 96. Alkylation of cysteine by iodoacetamide.

Cysteine alkylation with iodoacetamide is very specific at near-neutral pH (7 to 8.5) and is a well-known procedure in sample preparation for 2D gel electrophoresis [75, 76] to prevent formation of disulfides, or acrylamidation during migration within the gel; it has also been used to increase the information content of peptide mapping experiments [63], or to undergo quantitative comparison of two proteinaceous samples by the Isotope-Coded-Affinity-Tag (ICAT) approach [77-79], where iodoacetamide is linked to a biotin to allow isolation of labelled tryptic digested peptides. Here, the purpose was simply to count the total number of cysteines in β-lactoglobulin A by fully denaturing in 8M urea followed by DTT reduction and alkylation by iodoacetamide. Figure 97 shows the spectrum of β-lactoglobulin A infused from the microspray at a concentration of 2.7 µM in MeOH 50%: CH3COOH 1%: water (without any salt) at 200 nL/min. The molecular mass obtained from the spectrum is 18363.8 ± 6.1 Da.

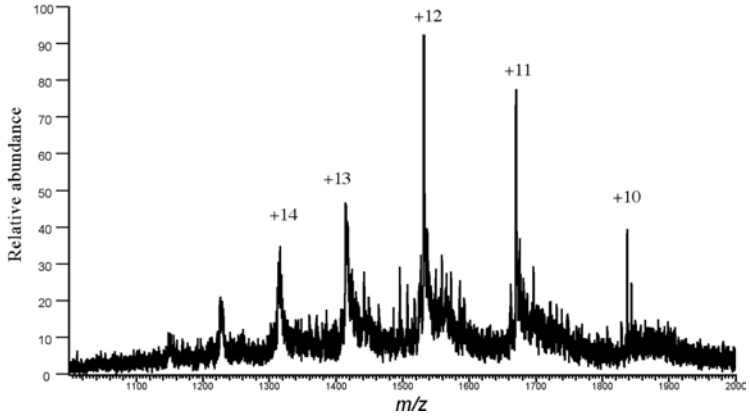


Figure 97. Spectrum of β-lactoglobulin A infused at 200 nL/min at a concentration of 2.7 μM; full-scale intensity is 2.25. 10^5 counts.

Figure 98 shows the spectrum of β-lactoglobulin A eluting from the membrane after denaturation, reduction by DTT and alkylation by iodoacetamide; desalting was done by pumping 50 μl of the incubated sample at 5 μL/min and washing out of salts with the same volume of water at the same flow rate. The resulting molecular mass is 18650.6 ± 2.9 Da. The mass shift corresponds exactly to the tagging of 5 cysteine moieties by iodoacetamide (Dm = 57 Da).

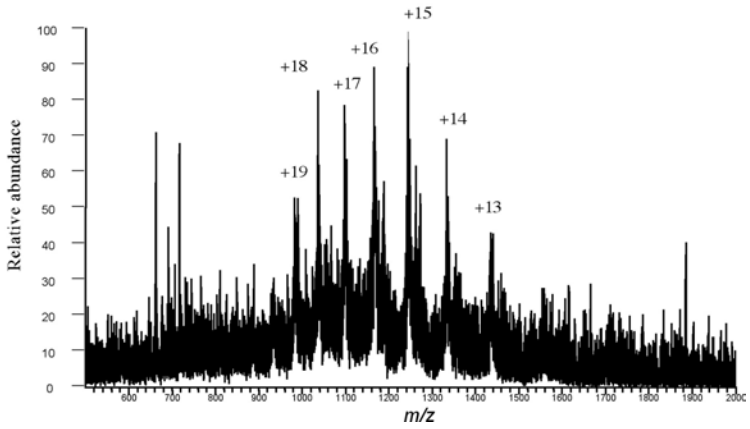


Figure 98. Spectrum of β-lactoglobulin A eluting from the PVDF membrane after denaturation, reduction, and alkylation with iodoacetamide.

5. Proteomic application of on-chip desalting

The results described in this section were obtained through a collaboration with the National High Magnetic Field Laboratory in Tallahassee, Florida, USA, thanks to a kind hosting by Alan Marshall and Mark Emmett, and with the technical assistance of Tukiet Lam and Logan Mackay.

5.1. Material and methods

Methanol (MeOH) and deionised water were from JT Baker (Phillipsburg, NJ, USA). Bovine ubiquitin (SwissProt accession number P02248, 76 amino acids, 8565 Da), horse heart myoglobin (SwissProt accession number P02188, 153 amino acids, 16951 Da), and formic acid were from Sigma (St. Louis, MO, USA). Pepsin was from Roche (Indianapolis, IN, USA). Poly(vinylidene difluoride) (PVDF) membranes were bought from Millipore (Immobilon P, Millipore, Bedford, MA, USA).

Microchips were kindly provided by DiagnoSwiss SA (Monthey, Switzerland). Microchip fabrication has been previously described [60, 61]. Copper-coated polyimide foils are subjected to several steps of chemical and plasma etching, resulting in 120 µm wide, 45 µm deep and 1 cm long microchannels. 52 µm in diameter gold microelectrodes sit at the bottom of the microchannel, one of which is used to deliver the electrospray voltage. The microchips are laminated with a polyethylene/polyethylene terephthalate lamination foil, and cut into a tip shape, so that the microchannel is opened on the edge of the chip, as described by Gobry *et al* [58].

The microchip coupling has been described previously [58, 59]. In the present study, a homemade microchip holder with low dead-volume connections is used to couple the microchip to an infusion line. The sample is delivered by a Harvard Apparatus syringe pump at a flow rate between 0.25 and 2 µL/min [80]. The electrospray voltage is applied

directly on the chip through a contact pad connected to the microelectrodes sitting in the microchannel. Electrospray voltage is 2 kV, and the chip is positioned in front of the heated capillary inlet so that the spacing between the microchannel outlet and the heated capillary is 2-5 mm.

The desalting process is the same as presented previously [81] (see section 4). Basically, a 3 mm diameter piece of PVDF membrane (120 µm thick) is wetted in pure MeOH and sandwiched between the microchannel inlet and the capillary connection, so that the only dead volume is the membrane itself. Infusion of a colored solution shows that 1 mm^2 of the membrane area is accessible. 10% MeOH (v:v) was added to the sample to wet the membrane as the sample is loaded at typical flow rates of 10 to 20 µl/min. The sample is then washed with an equal volume of 0.5% formic acid in water, and eluted with a methanol:water:formic acid mixture (see below for details) at the same flow rate

All experiments were performed with a homebuilt 9.4 T FT-ICR mass spectrometer as described elsewhere [82, 83]. Electrospray was generated by applying 2 kV and 350 V to the gold microelectrode on the microchip and the tube lens, respectively. Desolvation was assisted by the use of a heated metal capillary operating in constant current mode at 4 A, and ions were continuously externally accumulated (~1.7 s total accumulation time) in the front octopole (1.5 MHz at 0.7mV). Bruce-o-pole setting of 1.2 ms at 30 V was used for improved ion extraction from external accumulation [84]. Ions were then transferred to the open cylindrical ICR cell where they were trapped, excited and detected. The time-domain ICR signal (2 Mword) was subjected to baseline zeroing, followed by Hanning apodisation and one zero-fill before Fourier transformation, magnitude calculation, and frequency to mass conversion [85, 86] based on external mass calibration. All mass spectra were acquired by the use of the MIDAS data acquisition system developed in-house [87, 88].

Myoglobin was subjected to digestion with pepsin: three different mixtures of myoglobin:pepsin were incubated on ice for two minutes in 20 µL of water acidified with 0.5% formic acid. The ratios of myoglobin to pepsin were 100 pmol:100 pmol, 10 pmol:10 pmol, and 500 fmol:2 pmol. MeOH was then added to 10%, and the solution applied to the chip.

5.2. Results

Infusion experiments

To demonstrate the ease of coupling of the microchip microelectrospray with the homebuilt electrospray interface, simple protein solutions were sprayed through the microspray emitter. A 5 µM solution of ubiquitin was prepared in 50% MeOH: 49.5% water:0.5 % formic acid, and infused at 1 µL/min. Figure 99 (left) shows a single-scan FT-ICR mass spectrum of the ubiquitin solution, showing isotopic resolution of the 7+ and 6+ charge states (inserts) at resolving powers of 96500 and 82700. A 5 µM myoglobin solution was infused in the same way, resulting in the spectrum shown in Figure 99 (right), at a 56400 resolving power for the 18+ charge state (insert).

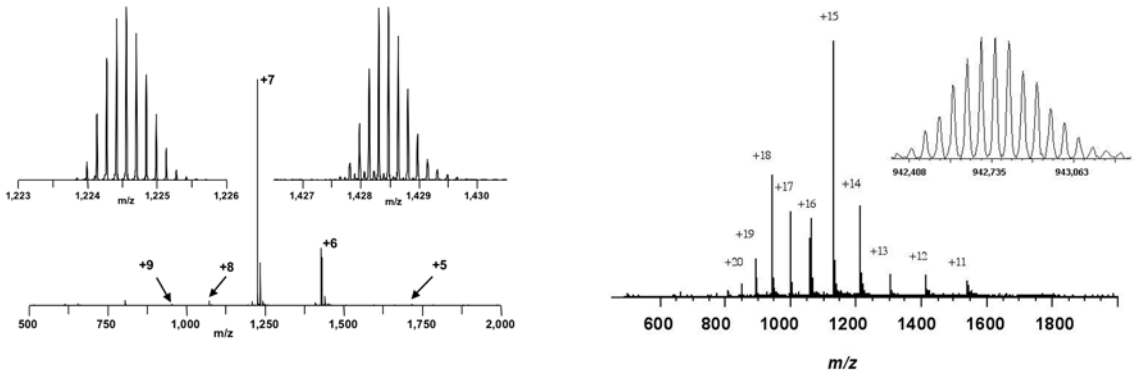


Figure 99. Left: FT-ICR single spectrum of 5 μM ubiquitin, with resolution of isotopic distributions for the 7+ and 6+ charge states (inserts) at resolving power of 96500 and 82700 respectively. Right: single spectrum of 5 μM myoglobin, with resolution of the isotopic distributions of the 18+ charge state (insert) at a resolving power of 56400.

Desalting of proteins

125 pmol of ubiquitin and 500 pmol of myoglobin were prepared in 20 μL of 1 M NaCl (resulting in 6.25 and 25 μM respectively). The two solutions were subjected to on-chip desalting as described above. Single-scan spectra obtained from the solution eluting from the PVDF membrane are shown in Figure 100, along with magnification of the isotopic resolution of the 7+ and 6+ charge states of ubiquitin and 18+ charge state of myoglobin (see inserts).

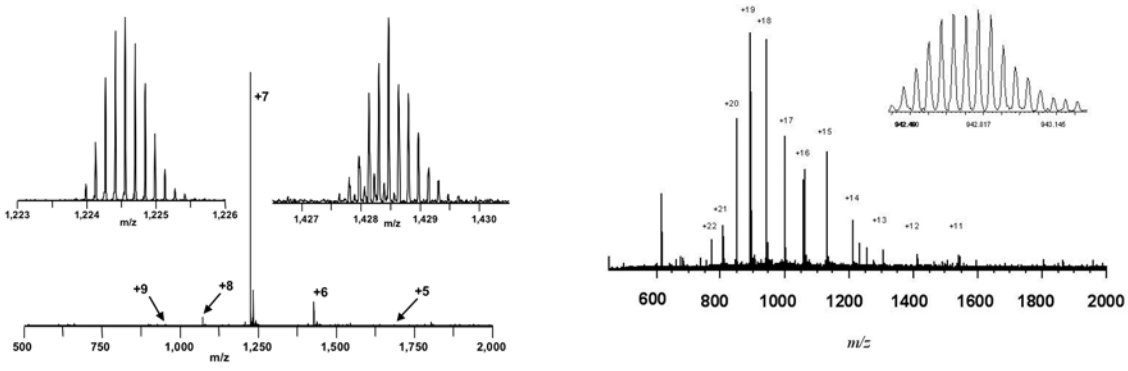


Figure 100. Single scan spectra of solutions eluting from the PVDF membrane. Left: 125 pmol of ubiquitin applied to the chip in 1 M NaCl, and Right: 500 pmol of myoglobin applied to the chip in 1 M NaCl.

The respective resolving powers are 128700 at *m*/*z*=1224 and 82700 at *m*/*z*=1428 for the two isotopic distributions of ubiquitin, and 49600 at *m*/*z*=942 for the 18+ charge state of myoglobin.

Desalting of peptides/pepsin digests

Determination of solvent accessibility in proteins and protein complexes by H/D exchange followed by enzymatic digestion and MS of dozens of the resulting peptides requires high sequence coverage as well as sufficient redundancy (defined herein as the number of peptides in which each amino acid is represented) [89]. Figure 101 shows 10 co-added ESI FT-ICR mass spectra of the peptides eluting from the membrane in 50% MeOH:49.5% water:0.5% formic acid following peptic digestion of 100 pmol of myoglobin.

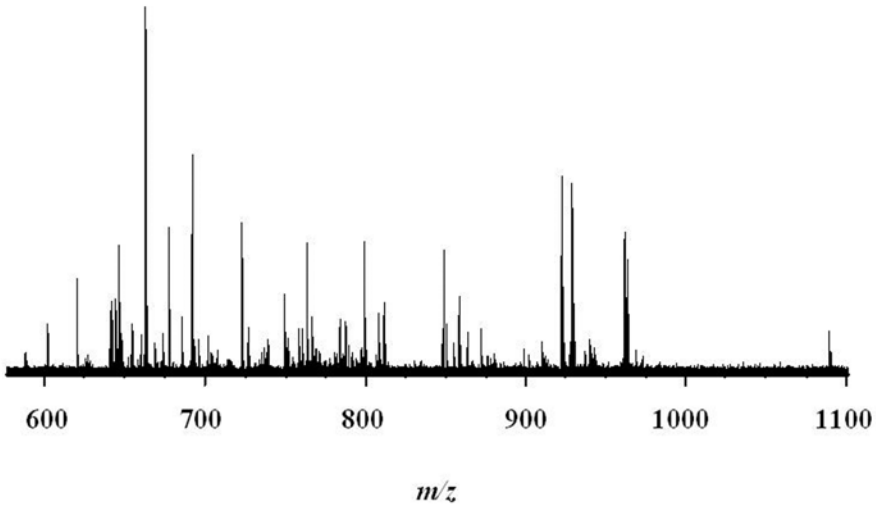


Figure 101. 10 scans co-added, corresponding to 100 pmoles digested myoglobin eluting from the chip in 50% MeOH: 49.5% water: 0.5% formic acid.

58 peptides can be identified from this single spectrum, resulting in 98.7% sequence coverage (only the two N-terminal amino acids are missing), with an average redundancy of 9.1 peptides per amino acid. The average mass accuracy is 9.9 ppm. At lower myoglobin concentration, digests were eluted from the membrane in 75% MeOH:24.5% water:0.5% formic acid in order to shorten the elution time and enhance sensitivity. Sequence coverage, mass accuracy, and redundancy for each of various experimental conditions are given in Table 17.

	Number of peptide matches	Sequence coverage (%)	Average mass accuracy (ppm)	Redundancy[4] (peptide/amino acid)
100 pmol	58	98.7	9.9 ± 7.7	9.1 ± 4.4
10 pmol (coadded)	16	96.7	19.0 ± 10.2	3.1 ± 2.0
500 fmol (coadded)	12	58.2	25.5 ± 11.7	2.1 ± 2.4

Table 17. Evaluation of myoglobin mapping by on-chip desalting of peptic digest.

[4] Redundancy gives the average number of peptides in which one particular amino acid is represented.

For the two lower concentrations (100 pmol and 500 fmol of starting material), the peptides elute differentially, so that it was necessary to co-add all spectra corresponding to the total elution period (10 spectra for the 100 pmol digest and 4 spectra for the 500 fmol digest) to obtain optimal sequence coverage. This difference in elution behaviour can be explained by the limited binding capacity of the membrane (85 µg/cm2, supplier data, or a binding capacity of ~100 pmol for the present membrane disk). The 100 pmol digest may thus be close to the saturation of the membrane, and consequently have a different elution behaviour. Several membranes can be stacked to increase the binding capacity if necessary, without a significant increase in back-pressure (data not shown).

However, for the lowest concentration tested here (500 fmol of myoglobin), the sequence coverage decreases to 60%, which may be satisfactory for protein identification by peptide mass fingerprinting, but not for protein mapping experiments. The minimal amount of starting material for this setup can thus be estimated to be between 10 pmol and 500 fmol to obtain good sequence coverage.

6. Conclusion

This chapter presented the integration of a solid-phase extraction step on microfabricated polymer microsprays. This new sample clean-up process can be used off-line, when capture membranes are incubated within the sample solutions to clean-up, and then placed on the microspray for subsequent elution, or directly used as a stop-and-go microspray with an infusion line. The sample is then pumped through the membrane which is further washed before application of the elution solution. The main features of this integrated sample clean-up process are its ease of use and quickness. The complete sample loading, washing and elution can be completed within 2-5 minutes. The main drawback of

this in-line process is that elution and spray onset have to be started at the same time in order to follow the whole elution process, which turns out to be difficult in practice with the prototype positioning devices used in this study: most of time, the chip positioning has to be adjusted after the beginning of elution to obtain good spraying performances. One possibility to circumvent this problem would be whether to develop a robust interface for reproducible and easy microspray positioning, or to use dual microsprays (as described in the following chapters): one microchannel would then be used to start the spray with a typical spraying solution, and the other microchannel would be used as described in this chapter for the SPE process. In this way, microspray positioning and electrospray onset could be guaranteed before starting the elution.

Whereas the proof-of-concept described in this chapter was conducted with PVDF membranes only, the same process can easily be adapted to membranes tailored with other functions and selectivities, such as antibody-grafted membranes for affinity capture and enrichment of specific compounds, which would prove useful in diagnostic applications where no monoclonal antibody is available or where analyte heterogeneity need to be characterised, even after antibody capture, or in proteomics applications where peptides need to be quantified after antibody capture [90]. Alternatively, membranes can come from an upstream process such as electroblotting after gel electrophoresis (even though elution of electroblotted proteins may prove challenging) or chemical processing [91-93].

7. References

[1] Tang, L., Kebarle, P., Effect of the conductivity of the electrosprayed solution on the electrospray current - factors determining analyte sensitivity in electrospray mass-spectrometry, *Anal. Chem.*, **1991**, 63(23): 2709-2715.

[2] Tang, L., Kebarle, P., Dependence of ion intensity in electrospray mass-spectrometry on the concentration of the analytes in the electrosprayed solution, *Anal. Chem.*, **1993**, 65(24): 3654-3668.
[3] Tang, L., *doctoral thesis*, University of Alberta, 1993.
[4] Iribarne, J. V., Thomson, B. A., Evaporation of small ions from charged droplets, *J. Chem. Phys.*, **1976**, 64(6): 2287-2294.
[5] Enke, C. G., A predictive model for matrix and analyte effects in electrospray ionization of singly-charged ionic analytes, *Anal. Chem.*, **1997**, 69(23): 4885-4893.
[6] Myland, J. C., Oldham, K. B., Overcoming electroneutrality: concentrative and electrical conditions inside a charged droplet of electrolyte solution, *J. Electroanal. Chem.*, **2002**, 522(2): 115-123.
[7] Myszka, D. G., He, X., Dembo, M., Morton, T. A., Goldstein, B., Extending the range of rate constants available from BIACORE: Interpreting mass transport-influenced binding data, *Biophys. J.*, **1998**, 75(2): 583-594.
[8] Schuck, P., Minton, A. P., Analysis of mass transport-limited binding kinetics in evanescent wave biosensors, *Anal. Biochem.*, **1996**, 240(2): 262-272.
[9] Kebarle, P., Tang, L., From ions in solution to ions in the gas-phase - the mechanism of electrospray mass-spectrometry, *Anal. Chem.*, **1993**, 65(22): 972A-986A.
[10] Constantopoulos, T. L., Jackson, G. S., Enke, C. G., Effects of salt concentration on analyte response using electrospray ionization mass spectrometry, *J. Am. Soc. Mass Spectrom.*, **1999**, 10(7): 625-634.
[11] Wilm, M., Mann, M., Analytical properties of the nanoelectrospray ion source, *Anal. Chem.*, **1996**, 68(1): 1-8.
[12] Cech, N. B., Enke, C. G., Practical implications of some recent studies in electrospray ionization fundamentals, *Mass Spectrom. Rev.*, **2001**, 20(6): 362-387.
[13] Loscertales, I. G., Delamora, J. F., Experiments on the Kinetics of Field Evaporation of Small Ions from Droplets, *J. Chem. Phys.*, **1995**, 103(12): 5041-5060.
[14] Labowsky, M., Fenn, J. B., de la Mora, J. F., A continuum model for ion evaporation from a drop: effect of curvature and charge on ion solvation energy, *Anal. Chim. Acta*, **2000**, 406(1): 105-118.
[15] Kebarle, P., A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry, *J. Mass Spectrom.*, **2000**, 35(7): 804-817.
[16] Gamero-Castano, M., de la Mora, J. F., Kinetics of small ion evaporation from the charge and mass distribution of multiply charged clusters in electrosprays, *J. Mass Spectrom.*, **2000**, 35(7): 790-803.
[17] Gamero-Castano, M., de la Mora, J. F., Direct measurement of ion evaporation kinetics from electrified liquid surfaces, *J. Chem. Phys.*, **2000**, 113(2): 815-832.
[18] de la Mora, J. F., Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism, *Anal. Chim. Acta*, **2000**, 406(1): 93-104.
[19] Felitsyn, N., Peschke, M., Kebarle, P., Origin and number of charges observed on multiply-protonated native proteins produced by ESI, *Int. J. Mass Spectrom.*, **2002**, 219(1): 39-62.
[20] Peschke, M., Blades, A., Kebarle, P., Charged states of proteins. Reactions of doubly protonated alkyldiamines with NH3: Solvation or deprotonation. Extension of two proton cases to multiply protonated globular proteins observed in the gas phase, *J. Am. Chem. Soc.*, **2002**, 124(38): 11519-11530.
[21] Schnier, P. D., Gross, D. S., Williams, E. R., Electrostatic Forces and Dielectric Polarizability of Multiply Protonated Gas-Phase Cytochrome-C Ions Probed by Ion/Molecule Chemistry, *J. Am. Chem. Soc.*, **1995**, 117(25): 6747-6757.

[22] Hautreux, M., Hue, N., de Kerdaniel, A. D., Zahir, A., Malec, V., Laprevote, O., Under non-denaturing solvent conditions, the mean charge state of a multiply charged protein ion formed by electrospray is linearly correlated with the macromolecular surface, *Int. J. Mass Spectrom.*, **2004**, 231(2-3): 131-137.
[23] Samalikova, M., Grandori, R., Role of opposite charges in protein electrospray ionization mass spectrometry, *J. Mass Spectrom.*, **2003**, 38(9): 941-947.
[24] Nesatyy, V. J., Suter, M. J. F., On the conformation-dependent neutralization theory and charging of individual proteins and their non-covalent complexes in the gas phase, *J. Mass Spectrom.*, **2004**, 39(1): 93-97.
[25] Znamenskiy, V., Marginean, I., Vertes, A., Solvated ion evaporation from charged water nanodroplets, *J. Phys. Chem. A*, **2003**, 107(38): 7406-7412.
[26] Cech, N. B., Enke, C. G., Effect of affinity for droplet surfaces on the fraction of analyte molecules charged during electrospray droplet fission, *Anal. Chem.*, **2001**, 73(19): 4632-4639.
[27] Gomez, A., Tang, K. Q., Charge and fission of droplets in electrostatic sprays, *Phys. Fluids*, **1994**, 6(1): 404-414.
[28] Ngoka, L. C. M., Gross, M. L., Novel sodium binding properties of some cyclopentapeptide endothelin a selective receptor antagonists: Electrospray and fast-atom-bombardment mass spectrometric studies, *Biochem. Biophys. Res. Commun.*, **1999**, 254(3): 713-719.
[29] Cooks, R. G., Patrick, J. S., Kotiaho, T., McLuckey, S. A., Thermochemical determinations by the kinetic method, *Mass Spectrom. Rev.*, **1994**, 13(4): 287-339.
[30] Feng, W. Y., Gronert, S., Lebrilla, C. B., Lithium and sodium ion binding energies of N-acetyl and N-glycyl amino acids, *J. Amer. Chem. Soc.*, **1999**, 121(6): 1365-1371.
[31] Lichtenberg, J., de Rooij, N. F., Verpoorte, E., Sample pretreatment on microfabricated devices, *Talanta*, **2002**, 56(2): 233-266.
[32] de Mello, A. J., Beard, N., Dealing with 'real' samples: sample pre-treatment in microfluidic systems, *Lab Chip*, **2003**, 3(1): 11N-19N.
[33] Lion, N., Rohner, T. C., Dayon, L., Arnaud, I. L., Damoc, E., Youhnovski, N., Wu, Z. Y., Roussel, C., Josserand, J., Jensen, H., Rossier, J. S., Przybylski, M., Girault, H. H., Microfluidic systems in proteomics, *Electrophoresis*, **2003**, 24(21): 3533-3562.
[34] Thurman, E. M., Mills, M. S., Solid-phase extraction: principles and practice, Wiley & Sons, New-York, 1998.
[35] Kutter, J. P., Jacobson, S. C., Ramsey, J. M., Solid phase extraction on microfluidic devices, *J. Microcolumn Sep.*, **2000**, 12(2): 93-97.
[36] Kutter, J. P., Jacobson, S. C., Matsubara, N., Ramsey, J. M., Solvent-programmed microchip open-channel electrochromatography, *Anal. Chem.*, **1998**, 70(15): 3291-3297.
[37] Li, J. J., LeRiche, T., Tremblay, T. L., Wang, C., Bonneil, E., Harrison, D. J., Thibault, P., Application of microfluidic devices to proteomics research - Identification of trace-level protein digests and affinity capture of target peptides, *Mol. Cell. Proteomics*, **2002**, 1(2): 157-168.
[38] Oleschuk, R. D., Shultz-Lockyear, L. L., Ning, Y., Harrison, D. J., Trapping of bead-based reagents within microfluidic systems: on-chip solid-phase extraction and electrochromatography, *Anal. Chem.*, **2000**, 72(3): 585-590.
[39] Jemere, A. B., Oleschuk, R. D., Ouchen, F., Fajuyigbe, F., Harrison, D. J., An integrated solid-phase extraction system for sub-picomolar detection, *Electrophoresis*, **2002**, 23(20): 3537-3544.

[40] Ekstrom, S., Malmstrom, J., Wallman, L., Lofgren, M., Nilsson, J., Laurell, T., Marko-Varga, G., On-chip microextraction for proteomic sample preparation of in-gel digests, *Proteomics*, **2002**, 2(4): 413-421.

[41] Bergkvist, J., Ekstrom, S., Wallman, L., Lofgren, M., Marko-Varga, G., Nilsson, J., Laurell, T., Improved chip design for integrated solid-phase microextraction in on-line proteomic sample preparation, *Proteomics*, **2002**, 2(4): 422-429.

[42] Lettieri, G. L., Dodge, A., Boer, G., de Rooij, N. F., Verpoorte, E., A novel microfluidic concept for bioanalysis using freely moving beads trapped in recirculating flows, *Lab Chip*, **2003**, 3(1): 34-39.

[43] Yu, C., Davey, M. H., Svec, F., Frechet, J. M. J., Monolithic porous polymer for on-chip solid-phase extraction and preconcentration prepared by photoinitiated in situ polymerization within a microfluidic device, *Anal. Chem.*, **2001**, 73(21): 5088-5096.

[44] Yu, C., Xu, M. C., Svec, F., Frechet, J. M. J., Preparation of monolithic polymers with controlled porous properties for microfluidic chip applications using photoinitiated free-radical polymerization, *J. Polym. Sci. Pol. Chem.*, **2002**, 40(6): 755-769.

[45] Liu, C. L., Hofstadler, S. A., Bresson, J. A., Udseth, H. R., Tsukuda, T., Smith, R. D., Snyder, A. P., On line dual microdialysis with ESI-MS for direct analysis of complex biological samples and microorganism lysates, *Anal. Chem.*, **1998**, 70(9): 1797-1801.

[46] Xu, N., Lin, Y., Hofstadler, S. A., Matson, D., Call, C. J., Smith, R. D., A microfabricated dialysis device for sample cleanup in electrospray ionization mass spectrometry, *Anal. Chem.*, **1998**, 70(17): 3553-2556.

[47] Xiang, F., Lin, Y., Wen, J., Matson, D. W., Smith, R. D., An integrated microfabricated device for dual microdialysis and on-line ESI-ion trap mass spectrometry for analysis of complex biological samples, *Anal. Chem.*, **1999**, 71(8): 1485-1490.

[48] Jiang, Y., Wang, P. C., Locascio, L. E., Lee, C. S., Integrated plastic microfluidic devices with ESI-MS for drug screening and residue analysis, *Anal. Chem.*, **2001**, 73(9): 2048-2053.

[49] Thornton, J. D., Science and practice of liquid-liquid extraction, Clarendon Press, Oxford, 1992.

[50] Peng, S. X., Branch, T. M., King, S. L., Fully automated 96-well liquid-liquid extraction for analysis of biological samples by liquid chromatography with tandem mass spectrometry, *Anal. Chem.*, **2001**, 73(3): 708-714.

[51] Yakhnin, A. V., Vinokurov, L. M., Surin, A. K., Alakhov, Y. B., Green fluorescent protein purification by organic extraction, *Protein Expr. Purif.*, **1998**, 14(3): 382-386.

[52] Weigl, B. H., Yager, P., Microfluidics - Microfluidic diffusion-based separation and detection, *Science*, **1999**, 283(5400): 346-347.

[53] Tokeshi, M., Minagawa, T., Kitamori, T., Integration of a microextraction system - Solvent extraction of a Co-2-nitroso-5-dimethylaminophenol complex on a microchip, *J. Chromatogr. A*, **2000**, 894(1-2): 19-23.

[54] Schilling, E. A., Kamholz, A. E., Yager, P., Cell lysis and protein extraction in a microfluidic device with detection by a fluorogenic enzyme assay, *Anal. Chem.*, **2002**, 74(8): 1798-1804.

[55] Ursitti, J. A., Mozdzanowski, J., Speicher, D. W., Electroblotting from polyacrylamide gels. In *Current Protocols in Protein Science*. Ed. Coligan, J. E., Dunn, B. M., Ploegh, H. L., Speicher, D. W., Wingfield, P. T. 1995. John Wiley & Sons, New York, USA, pp Unit 10.7.

[56] Magi, B., Bini, L., Marzocchi, B., Liberatori, S., Raggiaschi, R., Pallini, V., Blotting and immunoaffinity identification of two-dimensional electrophoresis-separated proteins. In *Proteome research: two-dimensional gel electrophoresis and identification methods*. Ed. Rabilloud, T. 2000. Springer-Verlag, Berlin Heidelberg New York, pp chapter 6 (pp 127-142).

[57] Millipore, Protein blotting handbook, *available from http://www.millipore.com*, **2002**, .

[58] Gobry, V., van Oostrum, J., Martinelli, M., Rohner, T., Rossier, J. S., Girault, H. H., Microfabricated polymer injector for direct mass spectrometry coupling, *Proteomics*, **2002**, 2(4): 405-412.

[59] Rohner, T. C., Rossier, J. S., Girault, H. H., Polymer Microspray with an Integrated Thick-Film Microelectrode, *Anal. Chem.*, **2001**, 73(22): 5353-5357.

[60] Rossier, J. S., Vollet, C., Carnal, A., Lagger, G., Gobry, V., Girault, H. H., Michel, P., Reymond, F., Plasma etched polymer microelectrochemical systems, *Lab Chip*, **2002**, 2(3): 145-150.

[61] Rossier, J., Reymond, F., Michel, P., Polymer microfluidic chips for electrochemical and biochemical analyses, *Electrophoresis*, **2002**, 23(6): 858-867.

[62] Lion, N., Gobry, V., Jensen, H., Rossier, J., Girault, H. H., Integration of a membrane-based desalting in a microfabricated disposable polymer injector for mass spectrometric protein analysis, *Electrophoresis*, **2002**, 23(20): 3583-3588.

[63] Sechi, S., Chait, B. T., Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification, *Anal. Chem.*, **1998**, 70(24): 5150-5158.

[64] Rabilloud, T., Chevallet, M., Solubilization of proteins in 2D electrophoresis. In *Proteome research: two-dimensional gel electrophoresis and identification methods*. Ed. Rabilloud, T. 2000. Springer-Verlag, Berlin Heidelberg New York, pp chapter 2 (pp 9-30).

[65] Babu, K. R., Douglas, D. J., Methanol-induced conformations of myoglobin at pH 4.0, *Biochemistry*, **2000**, 39(47): 14702-14710.

[66] Babu, K. R., Moradian, A., Douglas, D. J., The methanol-induced conformational transitions of beta- lactoglobulin, cytochrome c, and ubiquitin at low pH: A study by electrospray ionization mass spectrometry, *J. Am. Soc. Mass Spectrom.*, **2001**, 12(3): 317-328.

[67] Cunsolo, V., Foti, S., La Rosa, C., Saletti, R., Canters, G. W., Verbeet, M. P., Free energy for blue copper protein unfolding determined by electrospray ionisation mass spectrometry, *Rapid Commun. Mass Spectrom.*, **2001**, 15(19): 1817-1825.

[68] Kashiwagi, T., Yamada, N., Hirayama, K., Suzuki, C., Kashiwagi, Y., Tsuchiya, F., Arata, Y., Kunishima, N., Morikawa, K., An electrospray-ionization mass spectrometry analysis of the pH-dependent dissociation and denaturation processes of a heterodimeric protein, *J. Am. Soc. Mass Spectrom.*, **2000**, 11(1): 54-61.

[69] Grandori, R., Matecko, I., Muller, N., Uncoupled analysis of secondary and tertiary protein structure by circular dichroism and electrospray ionization mass spectrometry, *J. Mass Spectrom.*, **2002**, 37(2): 191-196.

[70] Grandori, R., Origin of the conformation dependence of protein charge-state distributions in electrospray ionization mass spectrometry, *J. Mass Spectrom.*, **2003**, 38(1): 11-15.

[71] Samalikova, M., Matecko, I., Muller, N., Grandori, R., Interpreting conformational effects in protein nano-ESI-MS spectra, *Anal. Bioanal. Chem.*, **2004**, 378(4): 1112-1123.

[72] Winston, R. L., Fitzgerald, M. C., Mass spectrometry as a readout of protein structure and function, *Mass Spectrom. Rev.*, **1997**, 16(4): 165-179.

[73] Kaltashov, I. A., Eyles, S. J., Studies of biomolecular conformations and conformational dynamics by mass spectrometry, *Mass Spectrom. Rev.*, **2002**, 21(1): 37-71.

[74] Herbert, B., Advances in protein solubilisation for two-dimensional electrophoresis, *Electrophoresis*, **1999**, 20(4-5): 660-663.

[75] Herbert, B., Galvani, M., Hamdan, M., Olivieri, E., MacCarthy, J., Pedersen, S., Righetti, P. G., Reduction and alkylation of proteins in preparation of two-dimensional map analysis: Why, when, and how?, *Electrophoresis*, **2001**, 22(10): 2046-2057.

[76] Galvani, M., Hamdan, M., Herbert, B., Righetti, P. G., Alkylation kinetics of proteins in preparation for two- dimensional maps: A matrix assisted laser desorption/ionization-mass spectrometry investigation, *Electrophoresis*, **2001**, 22(10): 2058-2065.

[77] Han, D. K., Eng, J., Zhou, H. L., Aebersold, R., Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry, *Nat. Biotechnol.*, **2001**, 19(10): 946-951.

[78] Zhou, H. L., Ranish, J. A., Watts, J. D., Aebersold, R., Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry, *Nat. Biotechnol.*, **2002**, 20(5): 512-515.

[79] Smolka, M. B., Zhou, H. L., Purkayastha, S., Aebersold, R., Optimization of the isotope-coded affinity tag-labeling procedure for quantitative proteome analysis, *Anal. Biochem.*, **2001**, 297(1): 25-31.

[80] Lion, N., Gellon, J. O., Girault, H. H., Flow rate characterization of microfabricated microspray emitters, *Rapid Comm. Mass Spectrom.*, **2004**, 18(14): 1614-1620.

[81] Lion, N., Gellon, J. O., Jensen, H., Girault, H. H., On-chip protein sample desalting and preparation for direct-coupling with electrospray ionization mass spectrometry, *J. Chromatogr. A*, **2003**, 1003(1-2): 11-19.

[82] Senko, M. W., Hendrickson, C. L., PasaTolic, L., Marto, J. A., White, F. M., Guan, S. H., Marshall, A. G., Electrospray ionization Fourier transform ion cyclotron resonance at 9.4 T, *Rapid Commun. Mass Spectrom.*, **1996**, 10(14): 1824-1828.

[83] Senko, M. W., Hendrickson, C. L., Emmett, M. R., Shi, S. D. H., Marshall, A. G., External accumulation of ions for enhanced electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry, *J. Am. Soc. Mass Spectrom.*, **1997**, 8(9): 970-976.

[84] Wilcox, B. E., Hendrickson, C. L., Marshall, A. G., Improved ion extraction from a linear octopole ion trap: SIMION analysis and experimental demonstration, *J. Am. Soc. Mass Spectrom.*, **2002**, 13(11): 1304-1312.

[85] Ledford, E. B., Rempel, D. L., Gross, M. L., Space-Charge Effects in Fourier-Transform Mass-Spectrometry - Mass Calibration, *Anal. Chem.*, **1984**, 56(14): 2744-2748.

[86] Shi, S. D. H., Drader, J. J., Freitas, M. A., Hendrickson, C. L., Marshall, A. G., Comparison and interconversion of the two most common frequency-to-mass calibration functions for Fourier transform ion cyclotron resonance mass spectrometry, *Int. J. Mass Spectrom.*, **2000**, 195-196(1): 591-598.

[87] Senko, M. W., Canterbury, J. D., Guan, S. H., Marshall, A. G., A high-performance modular data system for Fourier transform ion cyclotron resonance mass spectrometry, *Rapid Commun. Mass Spectrom.*, **1996**, 10(14): 1839-1844.

[88] Blakney, G. T., Hendrickson, C. L., Emmett, M. E., Marshall, A. G., Improved MIDAS data station for FT-ICR mass spectrometry, in *Proc. 50th Amer. Soc. Mass Spectrom. Conf.: Mass Spectrometry & Allied Topics*, 2002.

[89] Engen, J. R., Smith, D. L., Investigating protein structure and dynamics by hydrogen exchange MS, *Anal. Chem.*, **2001**, 73(9): 256A-265A.

[90] Anderson, N. L., Anderson, N. G., Haines, L. R., Hardie, D. B., Olafson, R. W., Pearson, T. W., Mass spectrometric quantitation of peptides and proteins using stable isotope standards and capture by anti-peptide antibodies (SISCAPA), *J. Proteome Res.*, **2004**, 3(2): 235-244.

[91] Lopez, M. F., Pluskal, M. G., Protein micro- and macroarrays: digitizing the proteome, *J. Chromatogr. B*, **2003**, 787(1): 19-27.

[92] Sloane, A. J., Duff, J. L., Wilson, N. L., Gandhi, P. S., Hill, C. J., Hopwood, F. G., Smith, P. E., Thomas, M. L., Cole, R. A., Packer, N. H., Breen, E. J., Cooley, P. W., Wallace, D. B., Williams, K. L., Gooley, A. A., High throughput peptide mass fingerprinting and protein macroarray analysis using chemical printing strategies, *Mol. Cell. Proteomics*, **2002**, 1(7): 490-499.

[93] Schulz, B. L., Packer, N. H., Karlsson, N. G., Small-scale analysis of O-linked oligosaccharides from glycoproteins and mucins separated by gel electrophoresis, *Anal. Chem.*, **2002**, 74(23): 6088-6097.

Chapter 6.Whole protein analysis by dual channel microsprayers

1. Introduction

Electrospray mass spectrometry is mainly used in proteomics for peptide analysis: when combined with proteolysis, the measurement of peptide masses and partial sequences by tandem mass spectrometry allow protein identification on a large scale. Whole protein analysis by electrospray mass spectrometry has been used traditionally for protein characterisation in various cases, such as for the control of the mass of a recombinant protein, the investigation of protein interactions after chemical cross-linking… But whole protein analysis does not allow protein identification, requires purified proteins, and usually results in poorly localised information along the protein sequence (such as the localisation of post-translational modifications).

One particular domain into which new breakthroughs have been obtained in the recent years is the analysis of very large multi-protein complexes: the group of Carol Robinson has been able to analyse intact ribosome complexes and determine their components and their stoechiometry [1, 2], characterise antibody-antigen complexes [3] or molecular chaperone complexes [4-7]. All these results proceed through analysis of gas phase ions in the range on megadaltons that has been made possible by developments of dedicated mass analysers.

Recently, new mass spectrometric procedures (collectively named top-down proteomics since they proceed from whole proteins) have been developed; basically they all rely on new high energy dissociation techniques that are able to fragment proteins in the gas phase, such as optimised collision-induced-dissociation (CID) [8], or electron capture dissociation (ECD) [9-11]. Basically, proteins are first separated in the liquid phase by reversed phase liquid chromatography or capillary electrophoresis, and then delivered to the mass spectrometer by electrospray ionisation. First intact protein masses are measured, and proteins are isolated in the gas phase, excited so that they fragment and fragment masses are measured, in which case the isolated protein can be identified by peptide mass fingerprinting, and its post-translational modifications characterised [12-14]. Top down proteomics can now be performed directly from bacterial lysates, and provides full characterisation of proteins up to a few tens of kilodaltons. Further technical developments in terms of ion accumulation, speed of ion processing in the gas phase, and sensitivity are necessary to enable the analysis of more complex proteomes. In addition, the range of molecular weights that are currently accessible to full gas phase fractionation is still limited to a small portion of proteomes. Nevertheless, top-down proteomics provides an analytical route alternative to shotgun proteomics. Similarly to 2D-GE combined with MALDI-MS, it does not necessitate complex database interrogation, as proteins are analysed one (or a few) at a time.

In this chapter, a new microfabricated microsprayer is introduced, which comprises two microchannels that have outlets in close proximity, on the edge of the microchip, so that the solutions contained in the two microchannels mix only within the Taylor cone, when electrospray is generated. By tuning solvent compositions and flow rates in the two microchannels, proteins can be sprayed in their native state, or denaturated within the Taylor cone by organic solvents or acids. Protein unfolding can thus be induced in solution

and measured from the mass spectrum through the charge state distribution of the protein in the mass spectrum.

2. Whole protein analysis by ESI-MS

2.1. Of the number of charges carried by a protein in vacuum

Protein analysis by mass spectrometry has been enabled by the development of soft ionisation sources, namely electrospray ionisation (ESI) and Matrix Assisted Laser Desorption Ionisation (MALDI) in the late eighties, that allow transfer of intact biomolecules from the liquid (ESI) or solid (MALDI) phase to the gas phase where they are separated according to their mass to charge ratio. A typical spectrum of a protein generated by electrospray mass spectrometry is shown Figure 102:

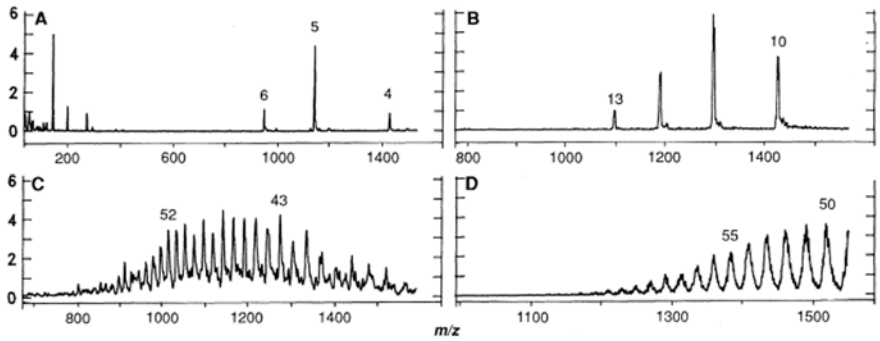


Figure 102. Electrospray mass spectra of (A) insulin (5730 Da), (B) lysozyme (14300 Da), (C) α-amylase (54700 Da) and (D) conalbumin (76000 Da). Reprinted with permission from [15].

Spectra display several peaks corresponding to different charge states of the protein; each peak thus has an *m*/*z* ratio of [M+nH]/n where M is the protein molecular weight, H the mass of a proton, and n the number of charges carried by the protein. The rationale for the number of charges carried by a protein transferred to vacuum by ESI as well as the distribution of *m*/*z* ratio has been the matter of a hot debate; a number of parameters of been

identified to influence the observed number of charges, such as the number of basic residues: arginine, lysine, histidine and the N terminus (in positive ion mode ESI) combined with the solvent used, the folding state of the protein, as well as instrumental parameters such as acceleration voltages or pressure of cooling gas used within the mass spectrometer. It is well established that an upper limit for the number of charges carried by a denatured protein in vacuum is the number of basic residues [16]:

$$protein - NH_2 + H_3O^+ \longrightarrow protein - NH_3^+ + H_2O$$

However, the mass spectrum of a protein displays a statistical bell-shape charge state distribution (CSD) whose mean value is not straightforwardly understandable from liquid-phase chemistry: just to name an example, Wang and Cole found that the ratio $[M+2H]^{2+}/[M+H]^+$ do not change in mass spectra of gramicidin S and bradykinin when the solution pH is varied from 3 to 11, whereas the acid-base chemistry predicts a change in this ratio from 10^5 to 10^{-2} for gramicidin S and 10^6 to 10^{-2} for bradykinin. Moreover, the same authors found that the ionic strength of the electrosprayed solution has no effect on protein charge states (as for example, myoglobin mean charge state changes from 14.46 to 14.05 when NH_4OAc concentration is increased from 0 to 0.3 M) [17]. Many other studies confirmed that neither solution acid-base chemistry nor the ionic strength have impact on the charge state distributions of proteins in ESI-MS. It is thus obvious that gas-phase processes play a major role on the number of charges carried by proteins in the gas phase. In summary, charged droplets are emitted from the apex of the Taylor cone and undergo coulombic fissions and solvent evaporation. At a certain point, small analytes can escape from electrospray droplets through columbic repulsion (charge evaporation model); alternatively, large analytes stay within the droplets until there is only one analyte molecule in the charged droplet surrounded by cations (charge residue model) (for a more detailed description of the electrospray process, see Chapter 5). Protein charge in the gas phase thus

depends on their ability to retain charges (protons in positive ion mode) within the electrospray droplets through columbic fissions and solvent evaporation. This affinity is described by the protein gas phase basicity:

$$M + H^+ \longleftrightarrow MH^+ \qquad GB(M) = -\Delta G_1 \qquad (1)$$

The case of charge ions is more complex, because of the presence of an activation barrier due to coulombic interactions:

$$MH^+ + H^+ \longleftrightarrow MH_2^{2+} \qquad GB(MH^+) = -\Delta G_2 \qquad (2)$$

Schnier *et al* have proposed a model for gas phase basicity of multiply charged ions [18]:

$$GB^{app} = GB_{\text{int},t} - \sum_{i=1}^{n} \frac{q^2}{(4\pi\varepsilon_0)\varepsilon_r r_{i,t}} \qquad (3)$$

Where GB^{app} is the apparent gas phase basicity of a protonation site *t* is the difference between its intrinsic gas phase basicity ($GB_{int,t}$) and the coulombic interaction experienced by this site from other protons. Authors further calculated the lowest energy proton configurations for each charge state using a pseudo random walk algorithm; from the calculated configurations, the gas phase basicity of each protonation site was calculated, and the global gas phase basicity of the charged protein was assigned the lowest gas phase basicity among all the protonation sites. The latest step was to compare the calculated gas phase basicities of multiply charged proteins within the gas phase basicity of the solvent in order to determine the highest charge state of a protein ion that would not undergo proton transfer to the solvent in the gas phase.

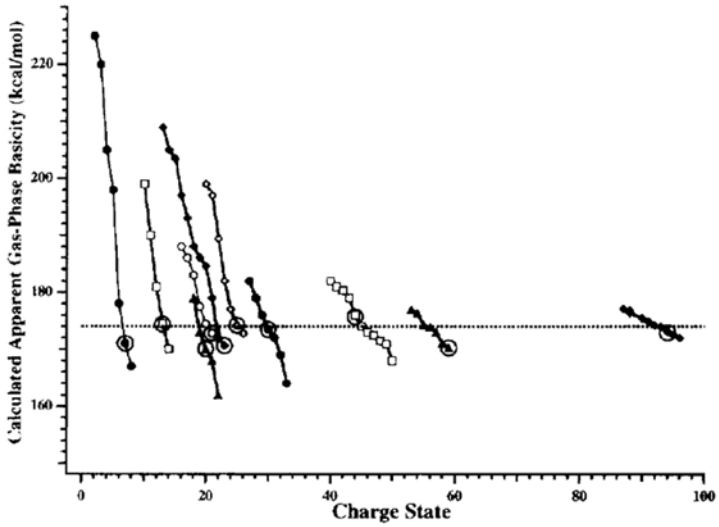


Figure 103. Calculated apparent gas phase basicities as a function of charge state for nine proteins. From left to right: melittin, ubiquitin, cytochrome C (tuna), cytochrome C (equine), ribonuclease A, hemoglobin (α chain), myoglobin, carbonic anhydrase, actin, and bovine serum albumin. The horizontal dashed line correspond to the gas phase of methanol (174.1 kcal/mol); circled values correspond to the maximal experimental charge states. Reprinted with permission from [18].

Figure 103 shows the excellent agreement between these calculations and the experimentally observed maximum charge state (circled values), which is often much lower than the number of basic residues. However in some cases, charge states higher than the number of basic residues are observed, such as for ribonuclease A (19 basic residues, 23 charges observed) or carbonic anhydrase (37 basic residues, 44 charges observed); in these cases, proline, glutamine and tryptophan are believed to retain additional protons. In summary, the number of basic residues provides an upper limit for the number of charges carried by a protein in the gas phase in most cases. However, the actual maximum charge state of a protein depends on its gas phase basicity compared to that of the solvent used for electrospray. In this case, calculations such as Eq. (3) provide a good estimate of this maximum number of charge.

2.2. Effect of conformation on charge state distributions of proteins

The calculations and experiments described above were all carried out with denatured proteins. However, it was early recognised that the folding state of a protein had an effect on its charge state distribution in ESI-MS [19]:

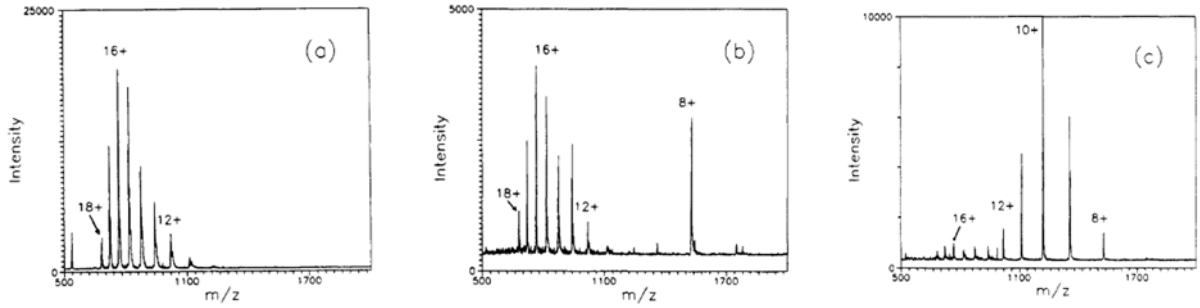


Figure 104. Electrospray mass spectra of cytochrome C (equine) at pH 2.6 (a), pH 3 (b) and pH 5.2 (c). Reprinted with permission from [19].

As shown in Figure 104, at acidic pH, cytochrome C exhibit a unimodal charge state distribution centred around the 15+ and 16+ peaks; at less acidic pH, a second distribution appears around the 8+ ion. At near neutral pH, the charge state distribution is centred around the 10+ ion. This shift in charge state distributions with acid concentration in methanol containing solutions has been lately correlated to cytochrome C folding states by circular dichroism [20], namely the full tertiary structure, the protein retaining the secondary structure with no tertiary structure, and the fully denatured protein.

There is now a clear qualitative consensus about the impact of protein conformation on charge state distribution in ESI-MS [21]: the more unfolded the protein, the more charges are retained during electrospray (and thus the lower the mass-to-charge ratio); the more folded the protein, the less charges are retained (and thus the higher the mass-to-charge ratio). However, the reasons for this effect are still highly debated. De La Mora proposed that native protein ions are formed via the charged residue model: proteins retain charges that would be present on a droplet of a similar charge at its Rayleigh limit [22].

However, this hypothesis was supported by very limited experimental sets, and did not provide much rationale to account for differences in mass spectra of folded and denatured proteins. That is the reason why Grandori recently proposed an alternative hypothesis: the reduction of charges observed on folded proteins compared to denatured ones would come from intramolecular neutralisation effects; protein charge states would thus simply be the difference between the numbers of acidic and basic sites [23]. Despite the decent agreement between this hypothesis and the experimental evidences presented [23, 24], the fact that protein charge state depends on intrinsic liquid phase parameters only (contrary to experimental evidences through years [17, 25-27]) rapidly cast doubts upon this hypothesis. More recently, Nesatyy and Suter re-established the charged residue model as the main mechanism for electrospray ionisation of folded proteins [28]:

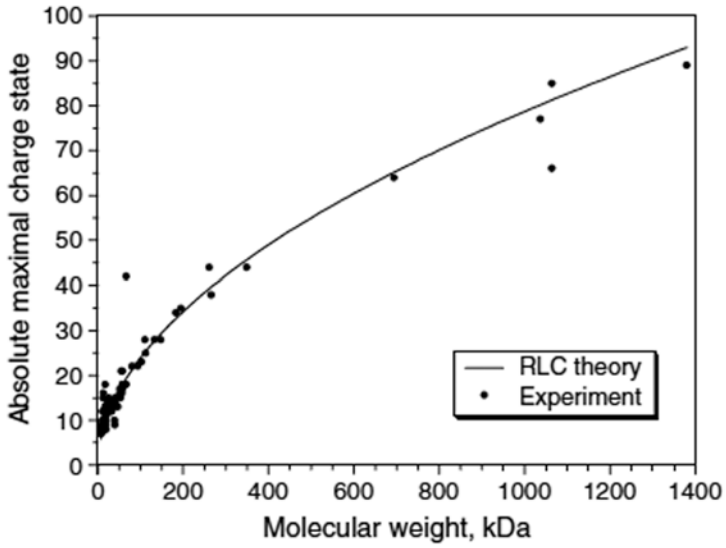


Figure 105. Correlation between the observed maximal charge state of 79 proteins and the charge carried by droplets of similar sizes at their Rayleigh limit (see Chapter 5 for more details). Reprinted with permission from [28].

Authors further established as a rule of thumb that the main charge state of a folded protein (most intense peak in the charge state distribution) is 0.85 of the calculated maximum charge state. The overall picture that begins to emerge is that charge state

distributions of proteins produced by electrospray do not depend on solution phase chemistry, but proceeds through the charge residue model, into which the protein charge is determined by the size of the electrosprayed protein: the number of charges carried by a protein in vacuum is the number of charges carried by a droplet of the same size at its Rayleigh limit. Differences between folded and unfolded proteins thus come solely from different sizes; however, if folded proteins can be approximated by charged spheres (and thus by charged droplets at their Rayleigh limit), unfolded proteins are usually assumed to adopt linear conformations whose correspondence with charged droplets is more difficult to delineate. In this prospect, the hyphenation of ion mobility analysers (IMA) with conventional time-of-flight (TOF) analysers which provides the measurement of collisional cross sections (IMA) together with charge state distributions (TOF) may allow to establish a direct correlation between protein conformation and charge state distribution (Carol Robinson, "Expanding Proteomics", Annual Meeting of the Swiss Proteomics Society, December 2005, 5-7, Zurich).

In this chapter a new microfabricated dual channel microsprayer is introduced for the analysis of intact proteins, with emphasis on the determination of protein conformation through the measurement of charge state distributions.

3. Materiel and Methods

3.1. Dual microsprayer fabrication

The dual channel microsprayer was fabricated according to the process described in Chapter 3 and kindly provided by DiagnoSwiss (Monthey, Switzerland). Basically, microchips consist in polyimide foils; one microchannel is etched on each side of the

microchip, so that they are superposed on the edge of the chip, where the electrospray is to be generated.

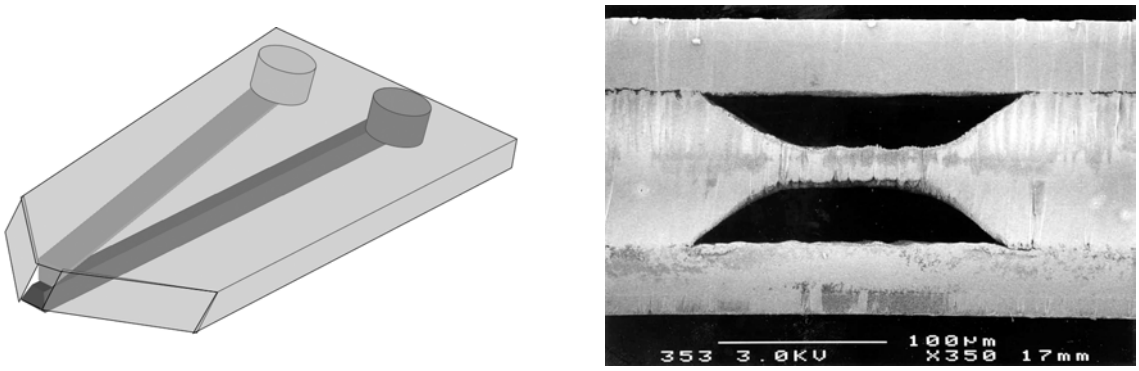


Figure 106. Left: scheme of the dual channel microsprayer with pasted reservoirs. Right: scanning electron microscopy picture of a cross section of the dual channel microsprayer. Microchannels are 120 µm wide, 45 µm deep, and are separated by 20 µm of polyimide.

Lastly, reservoirs are pasted over the inlets of the two microchannels in order to load samples and solvents.

3.2. Chemicals and reagents

Analytical grade methanol (MeOH) and trifluoroacetic acid (TFA) are from Riedel de Haen (Darmstaadt, Germany), acetic acid, propranolol-S-hydrochloride are from Fluka (Buchs, Switzerland), and HPLC grade acetonitrile (ACN) is from SDS (Peypin, France). Caffeine, reserpine, horse heart cytochrome C, hen egg white lysozyme and bovine myoglobin are from Sigma (St Louis, MO, USA). Deionised water (18.2 MW.cm) is obtained from a MilliQ unit (Millipore, Bedford, USA).

3.3. Mass spectrometry

Mass spectrometer is a LCQ Duo ion trap from Thermo Electron (San Jose, CA, USA). All experiments were done in full scan mode (50-2000 *m/z*) without any averaging, the heated capillary is kept at 200 °C, and ion optics parameters are optimized on the studied compound. The dual channel microsprayer is directly pasted on the x-y-z table with

double sided adhesive tape, and positioned in front of the heated capillary inlet, typically 3-5 mm away. The electrospray voltage is applied through a platinum wire plunge into one of the pasted reservoirs, and typically set between 2 and 3 kV. An additional voltage can be applied through a second platinum wire in the second reservoir thanks to a home made floating high power supply.

4. Results

4.1. Mode of operation

The dual channel microsprayer can be used under gravity induced flow alone or under a combination of gravity induced flow and electrokinetic flow (see Chapter 3). Given that two different voltages can be applied at two different points on the chip, the mode of operation is slightly more complex than a single channel microsprayer;

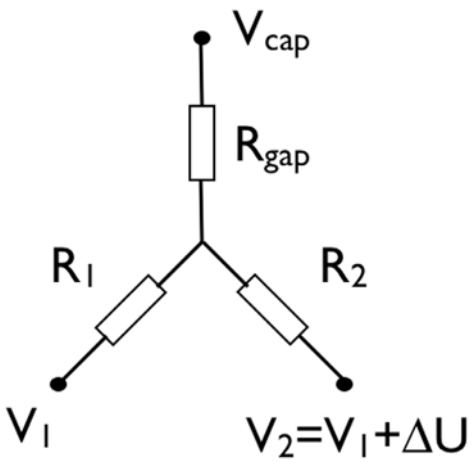


Figure 107. Electrical scheme of the dual channel microsprayer. V_1 and V_2 are the voltages applied in the two reservoirs, R_1 and R_2 the resistances of the two microchannels, R_{gap} the resistance of the air gap between the microchip and the mass spectrometer inlet (typically in the gigaohm range), and V_{cap} the voltage applied on the heated capillary (typically in the range –30 to +30 V). The node between the three branches represent the apex of the Taylor cone where electrospray is generated.

It is straightforward to calculate the voltage at the node between the three branches:

$$V_{tip} = \frac{\frac{V_{cap}}{R_{gap}} + \frac{V_1}{R_1} + \frac{V_2}{R_2}}{\frac{1}{R_{gap}} + \frac{1}{R_1} + \frac{1}{R_2}} \quad (4)$$

In most cases R_{gap} can be assumed to be large compared to R_1 and R_2, and V_{cap} to be small compared to V_1 and V_2. In this case, if the reservoir 1 is filled with an aqueous solution containing electrolytes and the reservoir 2 is filled with pure methanol (as is often the case in the experiments described below), then R_1 is small compared to R_2. It thus comes that V_{tip} is closed to V_1, and all ΔU is available to generate electroosmotic flow in microchannel 2, whereas there is only a negligible electric field to generate electroosmotic flow in microchannel 1.

4.2. Dual channel microsprayer characterisation

The initial purpose of this dual channel microsprayer was to tune the relative amounts of the two solutions (coming from the two microchannels) within the Taylor cone while spraying. In order to test the possibility to tune electrokinetically the flow rates in the two microchannels, 100 µM of propranolol and caffeine were loaded in pure water in one reservoir (20 µL), and 100 µM of reserpine is loaded in pure methanol in the other reservoir (20 µL). Electrospray voltage (3.2 kV) is applied in the reservoir containing reserpine in methanol, and additional voltage can be applied in the aqueous reservoir.

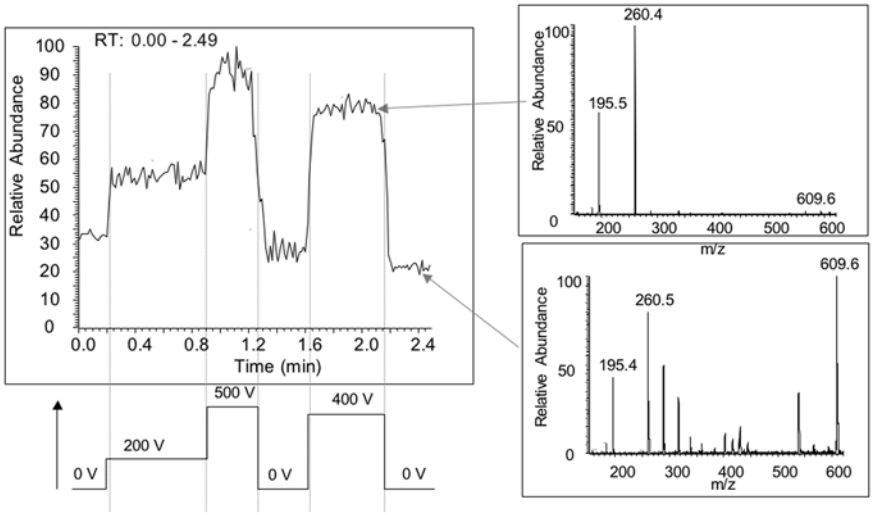


Figure 108. Characterisation of the dual channel microsprayer. Left: Total ion chromatogram with varying voltage applied in the aqueous reservoir (bottom). Right: resulting mass spectra at additional voltage 0V (bottom) and 400 V (top).

Figure 108 shows the effect of varying the additional voltage in the aqueous voltage: at 0 V, the total ion count is low and the mass spectrum contains caffein (m/z=195.4), propranolol (m/z=260.5), and reserpine (m/z=609.6), in comparative amounts. This means that both microchannels are operated under gravity induced flow, and all analytes are thus present in equal amounts in the mass spectrum. When the additional voltage in the aqueous microchannel is shifted to 400 V, the total ion count increases (approximately threefold), and reserpine almost disappears from the mass spectrum, due to the large increase in caffeine and propranolol intensities. This corresponds to a large increase in the flow rate coming from the aqueous microchannel due to the additional electroosmotic flow.

In order to properly measure the gain in sensitivity that can be obtained by tuning the relative flow rates between the organic sheath flow and the aqueous sample, 50 µM of reserpine in methanol was loaded in one reservoir (20 µL), and 100 µM of propranolol in water was loaded in the other reservoir (20 µL). Electrospray voltage was applied to the organic reservoir (3.0 kV), and the additional voltage (ΔU) to the aqueous reservoir.

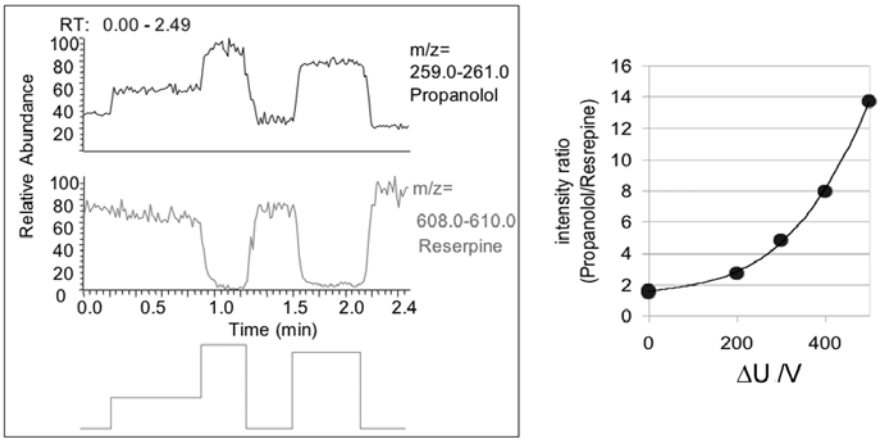


Figure 109. Analysis of aqueous propranolol with reserpine in the organic sheath flow.Left: selected ion chromatograms of propranolol and reserpine, and ΔU values. Right: ratio between propranolol and reserpine.

Figure 109 shows the effect of ΔU on the mass spectrum of propranolol and reserpine: for ΔU=0 V, both propranolol and reserpine are visible in the spectrum (intensity ratio around 2). When ΔU increases, more aqueous solution is electrokinetically pumped into the Taylor cone, which results in a large increase in propranolol intensity. Above ΔU=400 V, the spray became instable, most probably because higher flow rates necessitate higher electrospray voltage (in this experiment, electrospray voltage was kept at 3.0 kV).

These two experiments show that composition of the electrosprayed solution can be tuned by electrokinetically varying the flow rate in one of the microchannels. Doing so, it is possible to inject more aqueous solution containing the sample, and thus adjusting the relative amount of organic modifier, or alternatively control the amount of organic modifier injected in a constant flow of aqueous sample.

4.3. Analysis of intact protein

In order to explore the ability to spray with the dual channel microsprayer in various solvent conditions, 50 µM of cytochrome C was loaded in water with 40 mM of ammonium acetate and various amounts of acetic acid in one reservoir (20 µL), and 20 µL of methanol was loaded in the other reservoir. Electrospray voltage (2.8 kV) was applied in the aqueous reservoir.

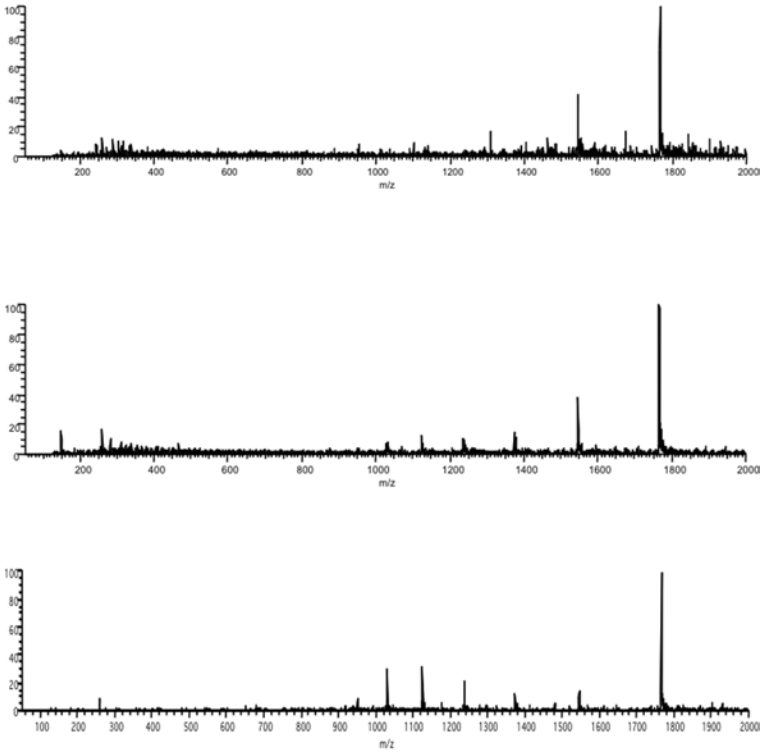


Figure 110. Acid induced denaturation of cytochrome C. Top: no acetic acid; middle: 0.1 % acetic acid; bottom: 1% of acetic acid.

As shown in Figure 110, when no acid is present, only the high *m/z* distribution of cytochrome C (8+ charge state) is visible, which correspond to native cytochrome C. Analysis of a purely aqueous solution of cytochrome C is completely impossible with a microchip embedding a single channel of similar dimensions. When acetic acid concentration increases, the unfolded distribution of cytochrome C, centred around 13+ charge state appears, which is consistent with the literature [20, 29]. In order to validate that

cytochrome C denaturation is due solely to the presence of acetic acid, and not to electrospray ionisation itself, cytochrome C was replaced by lysozyme, a protein that is known to be highly resistant to acid induced denaturation.

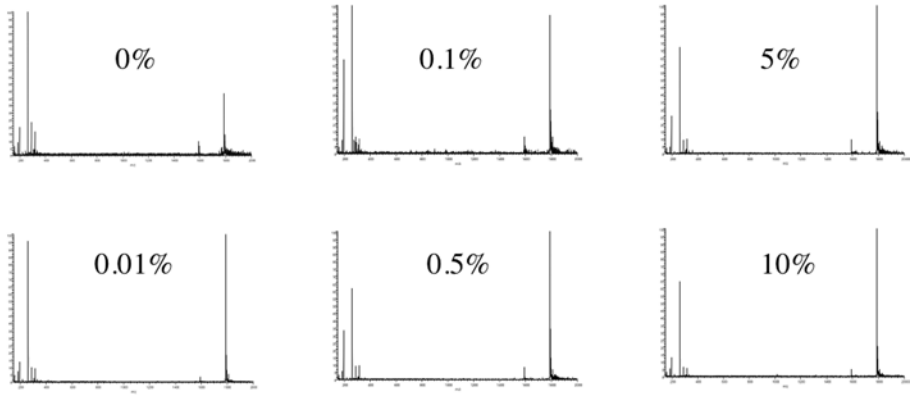


Figure 111. Analysis of 50 µM of lysozyme in 40 mM ammonium acetate and various amounts of acetic acid in the aqueous reservoir. Organic reservoir is pure methanol. Peaks in the low mass range are internal standards.

Figure 111 testifies that the electrospray process itself cannot account for cytochrome C denaturation in the experiment above, because lysozyme is always observed in its folded state, whatever the acetic acid amount added.

If the experiment is done in the other way, i.e. if the aqueous solution is always the same (50 µM of cytochrome C in 40 mM of ammonium acetate) and various amounts of acetic acid are added to the organic reservoir, the same kind of results is obtained:

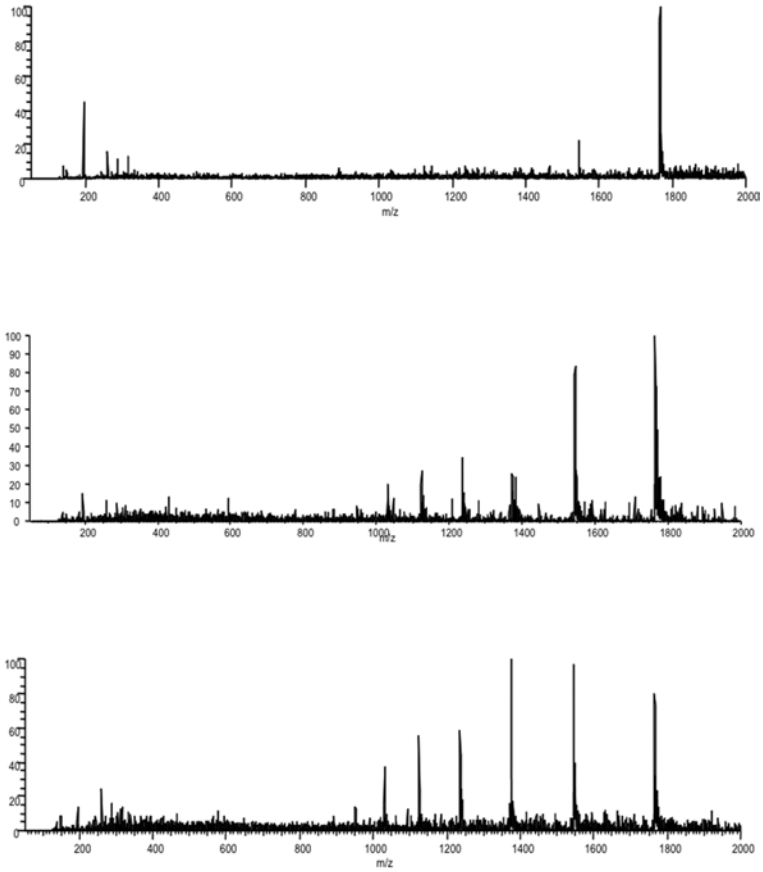


Figure 112. Cytochrome C denaturation within the Taylor cone. Acetic acid is added to the organic reservoir: 0% (top), 0.1% (middle) and 1% (bottom).

Again the native distribution of cytochrome C is observed when no acid is present, and the unfolded distribution appears when acetic acid content increases. The unique feature is here that protein denaturation takes place within the Taylor cone during electrospray ionisation: cytochrome C does not meet acetic acid before it reaches the Taylor cone.

The ability to denature proteins within the Taylor cone opens the possibility for online tuneable protein denaturation during electrospray ionisation: 50 µM of cytochrome C in 40 mM ammonium acetate was loaded in the aqueous reservoir, into which electrospray voltage was applied (3.2 kV). Pure methanol was loaded into the organic reservoir, into which an additional platinum electrode allow to apply an additional voltage.

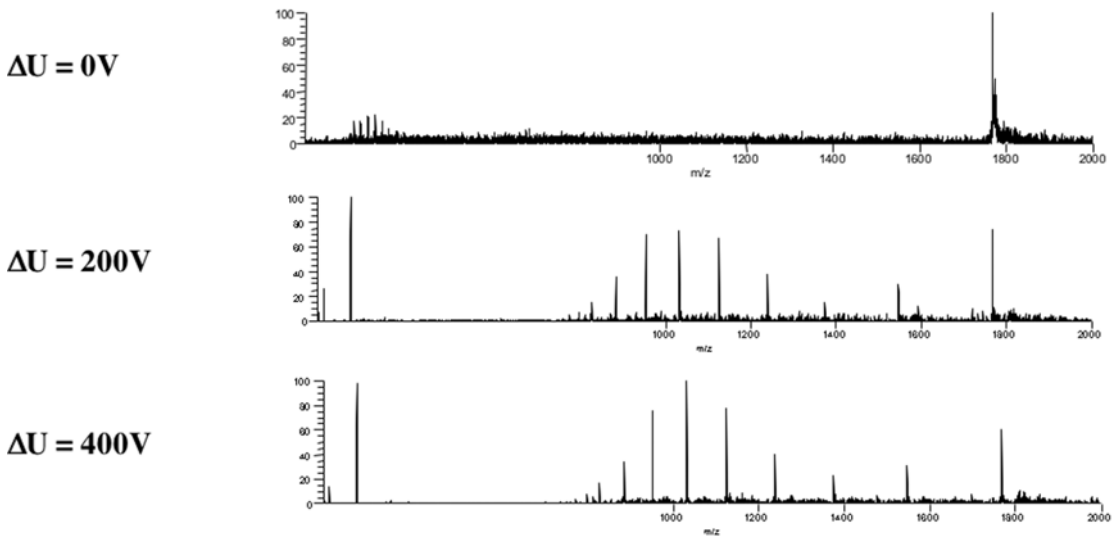


Figure 113. On-chip tuneable cytochrome C denaturation.

Figure 113 shows the effect of varying the additional voltage (ΔU) in the methanol reservoir on the spectrum: at ΔU=0V, minimum amount of methanol is incorporated in the spray, and cytochrome C remains in its native distribution. When ΔU increases, more methanol is pumped into the Taylor cone and induces denaturation of cytochrome C [20, 29].

Similarly, the possibility to analyse folded proteins pushes toward the analysis of non covalent protein complexes: the aqueous reservoir was loaded with holomyoglobin at 50 µM in 40 mM of ammonium acetate, and the organic reservoir with pure methanol. Electrospray voltage was applied to the aqueous reservoir (2.9 kV).

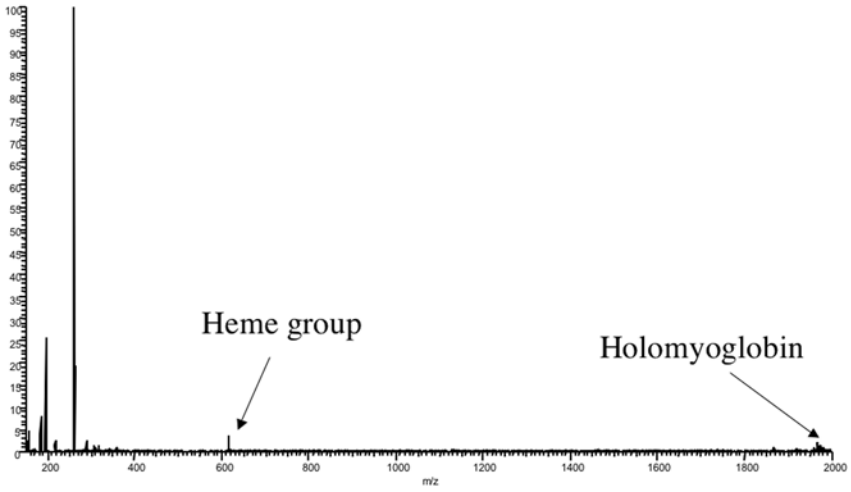


Figure 114. Analysis of holomyoglobin (low mass ions are internal standards).

Holomyoglobin exhibits very low intensity, because its isoelectric point is below the pH used here. The heme group appears alone, as is always the case in myoglobin samples. However, ion trap isolation of the ion present around m/z= 1950 allows to measure more precisely the mass-to-charge of the ion, which correspond exactly to the 9+ charge state of holomyoglobin.

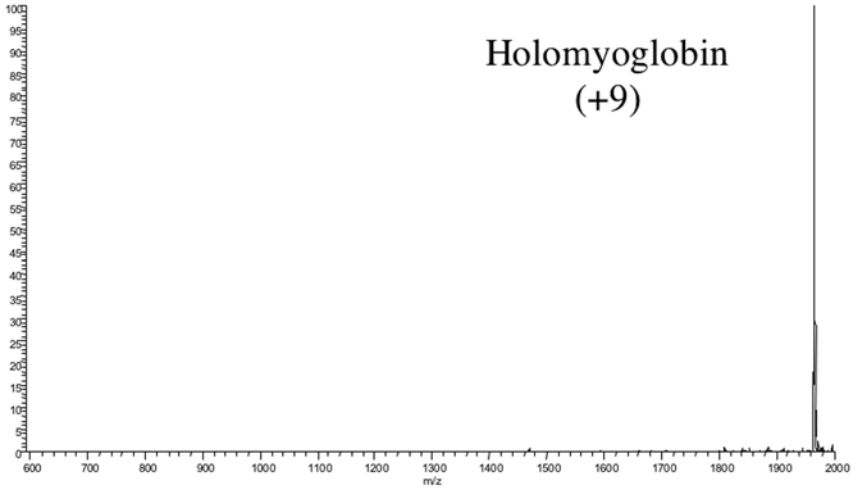


Figure 115. Ion trap isolation of the 9+ ion of holomyoglobin.

In order to further confirm the identity of this ion, it was subjected to low energy ion trap collision induced dissociation (CID at 20%):

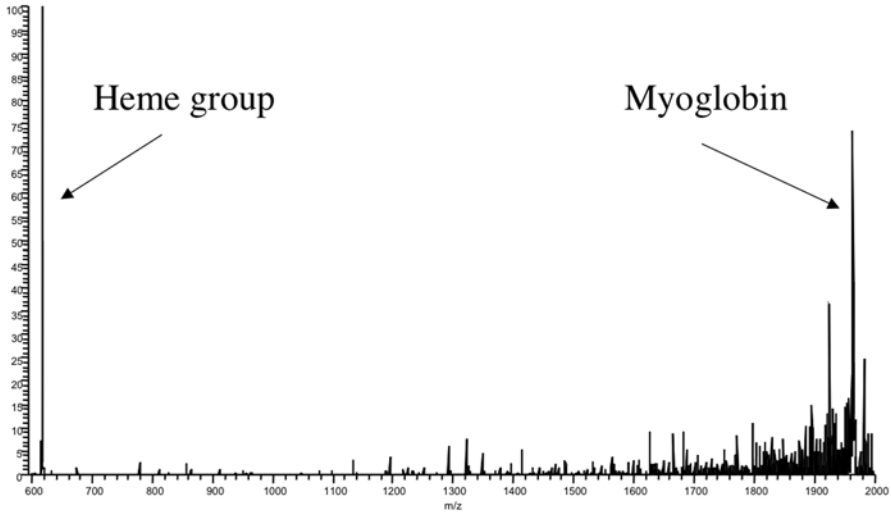


Figure 116. Low energy CID of holomyoglobin.

The low energy CID of the putative 9+ ion of holomyoglobin results in the release of heme (at m/z=616) and a major peak in the high mass range that corresponds to the 9+ ion of apomyoglobin.

5. Conclusion and perspectives

In this chapter, a new microfabricated dual channel microsprayer has been introduced for the analysis of intact proteins in the context of conformational analysis and for the analysis of non covalent complexes. Though this set of experiments has been limited by the fact that all experiments were performed on an ion trap instrument with a low mass range, it demonstrates the potential of this new microsprayer for folding studies. In particular, the on-chip tuneable denaturation of cytochrome C by methanol opens the way for full characterisation of protein denaturation: though the example shows only the spectra of cytochrome C at three different amount of methanol injected into the Taylor cone (see

Figure 113), it may be feasible to obtain more continuous values from intact cytochrome C (top row in Figure 113) to full denaturation of cytochrome C. From such an experiment, it would then be feasible to plot the average charge state of the protein versus ΔU (or ideally versus the flow rate of methanol). A comparable experiment has been performed by Grandori *et al* with multiple off-line experiments for the acid-induced denaturation of cytochrome C [30]:

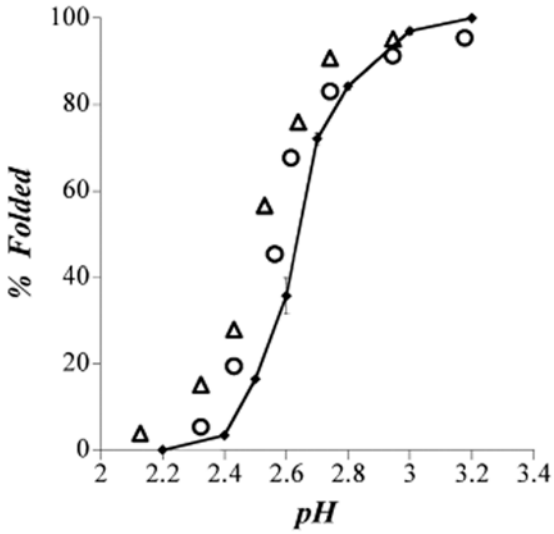


Figure 117. Acid-induced denaturation of cytochrome C measured by nanospray ESI-MS (full line), tryptophan fluorescence (open circles), and Soret absorption (open triangles).

As shown in Figure 117, acid-induced denaturation of cytochrome C as monitored by ESI-MS provides results in good agreement with other biophysical methods. In order to gain approval by the community, this methodology still needs to be tested on other proteins, side by side with reference biophysical methods, such as circular dichroism. If the generality of the method is proven, mass spectrometry may provide a breakthrough in routine thermodynamic studies of protein stability, in the measure that it can work with minute amounts of proteins (nanomole range used above compared to micromole range for most biophysical methods), with poorly purified proteins, contrary to optical methods, it can differentiate co-existing conformations by the existence of multimodal charge state

distributions in the mass spectrum, and does not rely on the presence of any particular probe within the protein structure (as is the case with tryptophan fluorescence).

More technical applications of such a new device could be for the hyphenation of liquid chromatography with ESI-MS: one microchannel could be connected to the HPLC eluent while the other could serve as a tuneable sheath flow for optimal ionisation of hydrophilic fractions. One other possible application is to perform post-column labelling: for example, the peptide KCTCCA was loaded in water with 40 mM of ammonium acetate and 1% acetic acid in one reservoir into which electrospray voltage was applied, and carboxymethylbenzoquinone in pure methanol in the other reservoir (in acidic conditions, carboxymethylbenzoquinone reacts selectively with cysteine residues [31, 32]):

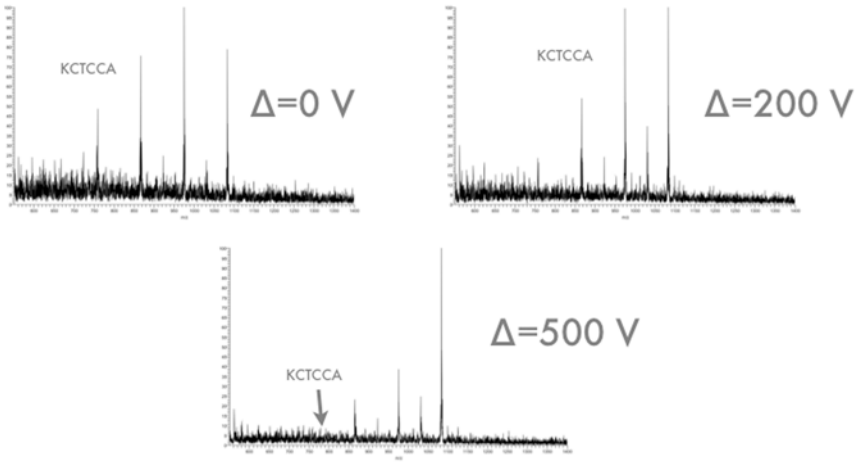


Figure 118. Post column labelling of the peptide KCTCCA with carboxymethylbenzoquinone.

Figure 118 shows the effect of the additional voltage applied in the organic reservoir to the tagging of the peptide KCTCCA: at ΔU=0V, there is enough carboxymethylhydroquinone pumped by gravity induced flow to tag the three cysteines, resulting in three additional peaks. When ΔU is increased to 200 V, more carboxymethylhydroquinone is pumped into the Taylor cone, resulting in a decrease of the

untagged KCTCCA peak, whereas adduct peaks increase. If ΔU is further increased, the untagged peptide peak disappears, giving rise to more intense adduct peaks. This simple experiment shows that the dual channel microsprayer can also be used for post-column labelling and tuned depending on the desired tagging effect.

6. References

[1] Benjamin, D. R., Robinson, C. V., Hendrick, J. P., Hartl, F. U., Dobson, C. M., Mass spectrometry of ribosomes and ribosomal subunits, *Proc. Natl. Acad. Sci. U. S. A.*, **1998**, 95(13): 7391-7395.

[2] Rostom, A. A., Fucini, P., Benjamin, D. R., Juenemann, R., Nierhaus, K. H., Hartl, F. U., Dobson, C. M., Robinson, C. V., Detection and selective dissociation of intact ribosomes in a mass spectrometer, *Proc. Natl. Acad. Sci. U. S. A.*, **2000**, 97(10): 5185-5190.

[3] Tito, M. A., Miller, J., Walker, N., Griffin, K. F., Williamson, E. D., Despeyroux-Hill, D., Titball, R. W., Robinson, C. V., Probing molecular interactions in intact antibody: Antigen complexes, an electrospray time-of-flight mass spectrometry approach, *Biophys. J.*, **2001**, 81(6): 3503-3509.

[4] Fandrich, M., Tito, M. A., Leroux, M. R., Rostom, A. A., Hartl, F. U., Dobson, C. M., Robinson, C. V., Observation of the noncovalent assembly and disassembly pathways of the chaperone complex MtGimC by mass spectrometry, *Proc. Natl. Acad. Sci. U. S. A.*, **2000**, 97(26): 14151-14155.

[5] Hernandez, H., Robinson, C. V., Dynamic protein complexes: Insights from mass spectrometry, *J. Biol. Chem.*, **2001**, 276(50): 46685-46688.

[6] Aquilina, A. J., Benesch, J. L. P., Bateman, O. A., Slingsby, C., Robinson, C. V., Polydispersity of a mammalian chaperone: Mass spectrometry reveals the population of oligomers in alpha B-crystallin, *Proc. Natl. Acad. Sci. U. S. A.*, **2003**, 100(19): 10611-10616.

[7] Sobott, F., Robinson, C. V., Characterising electrosprayed biomolecules using tandem-MS - the noncovalent GroEL chaperonin assembly, *Int. J. Mass Spectrom.*, **2004**, 236(1-3): 25-32.

[8] Keller, K. M., Brodbelt, J. S., Hettich, R. L., Van Berkel, G. J., Comparison of sustained off-resonance irradiation collisionally activated dissociation and multipole storage-assisted dissociation for top-down protein analysis, *J. Mass Spectrom.*, **2004**, 39(4): 402-411.

[9] Sze, S. K., Ge, Y., Oh, H., McLafferty, F. W., Top-down mass spectrometry of a 29-kDa protein for characterization of any posttranslational modification to within one residue, *Proc. Natl. Acad. Sci. U. S. A.*, **2002**, 99(4): 1774-1779.

[10] Patrie, S. M., Charlebois, J. P., Whipple, D., Kelleher, N. L., Hendrickson, C. L., Quinn, J. P., Marshall, A. G., Mukhopadhyay, B., Construction of a hybrid quadrupole/Fourier Transform Ion Cyclotron Resonance Mass Spectrometer for

versatile MS/MS above 10 kDa, *J. Am. Soc. Mass Spectrom.*, **2004**, 15(7): 1099-1108.

[11] Zubarev, R. A., Electron-capture dissociation tandem mass spectrometry, *Curr. Opin. Biotechnol.*, **2004**, 15(1): 12-16.

[12] Wu, S. L., Jardine, I., Hancock, W. S., Karger, B. L., A new and sensitive on-line liquid chromatography/mass spectrometric approach for top-down protein analysis: the comprehensive analysis of human growth hormone in an E. coli lysate using a hybrid linear ion trap/Fourier transform ion cyclotron resonance mass spectrometer, *Rapid Commun. Mass Spectrom.*, **2004**, 18(19): 2201-2207.

[13] Du, Y., Meng, F. Y., Patrie, S. M., Miller, L. M., Kelleher, N. L., Improved molecular weight-based processing of intact proteins for interrogation by quadrupole-enhanced FT MS/MS, *J. Proteome Res.*, **2004**, 3(4): 801-806.

[14] Patrie, S. M., Robinson, D. E., Meng, F. Y., Du, Y., Kelleher, N. L., Strategies for automating top-down protein analysis with Q-FTICR MS, *Int. J. Mass Spectrom.*, **2004**, 234(1-3): 175-184.

[15] Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., Whitehouse, C. M., Electrospray ionization for mass-spectrometry of large biomolecules, *Science*, **1989**, 246(4926): 64-71.

[16] Covey, T. R., Bonner, R. F., Shushan, B. I., Henion, J., *Rapid Commun. Mass. Spectrom.*, **1988**, 2(249-256.

[17] Wang, G., Cole, R. B., Effect of solution ionic strength on analyte charge state distributions in positive and negative electrospray mass spectrometry, *Anal.Chem.*, **1994**, 66(21): 3702-3708.

[18] Schnier, P. D., Gross, D. S., Williams, E. R., On the maximum charge-state and proton-transfer reactivity of peptide and protein ions formed by electrospray-ionization, *J. Am. Soc. Mass Spectrom.*, **1995**, 6(11): 1086-1097.

[19] Chowdhury, S. K., Katta, V., Chait, B. T., Probing conformational-changes in proteins by mass-spectrometry, *J. Am. Chem. Soc.*, **1990**, 112(24): 9012-9013.

[20] Konermann, L., Douglas, D. J., Acid-induced unfolding of cytochrome c at different methanol concentrations: Electrospray ionization mass spectrometry specifically monitors changes in the tertiary structure, *Biochemistry*, **1997**, 36(40): 12296-12302.

[21] Kaltashov, I. A., Eyles, S. J., Studies of biomolecular conformations and conformational dynamics by mass spectrometry, *Mass Spectrom. Rev.*, **2002**, 21(1): 37-71.

[22] de la Mora, J. F., Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism, *Anal. Chim. Acta*, **2000**, 406(1): 93-104.

[23] Grandori, R., Origin of the conformation dependence of protein charge-state distributions in electrospray ionization mass spectrometry, *J. Mass Spectrom.*, **2003**, 38(1): 11-15.

[24] Samalikova, M., Grandori, R., Role of opposite charges in protein electrospray ionization mass spectrometry, *J. Mass Spectrom.*, **2003**, 38(9): 941-947.

[25] Wang, G. D., Cole, R. B., Disparity between solution-phase equilibria and charge-state distributions in positive-ion electrospray mass-spectrometry, *Org. Mass Spectrom.*, **1994**, 29(8): 419-427.

[26] Wang, G. D., Cole, R. B., Effects of solvent and counterion on ion pairing and observed charge states of diquaternary ammonium salts in electrospray ionization mass spectrometry, *J. Am. Soc. Mass Spectrom.*, **1996**, 7(10): 1050-1058.

[27] Wang, G., Cole, R. B., Solution, gas-phase, and instrumental parameters influences on charge-state distributions in electrospray ionization mass spectrometry. In

Electrospray ionization mass spectrometry. Ed. Cole, R. B. 1997. John Wiley & Sons, New-York, pp 137-174.

[28] Nesatyy, V. J., Suter, M. J. F., On the conformation-dependent neutralization theory and charging of individual proteins and their non-covalent complexes in the gas phase, *J. Mass Spectrom.*, **2004**, 39(1): 93-97.

[29] Babu, K. R., Moradian, A., Douglas, D. J., The methanol-induced conformational transitions of beta- lactoglobulin, cytochrome c, and ubiquitin at low pH: A study by electrospray ionization mass spectrometry, *J. Am. Soc. Mass Spectrom.*, **2001**, 12(3): 317-328.

[30] Samalikova, M., Matecko, I., Muller, N., Grandori, R., Interpreting conformational effects in protein nano-ESI-MS spectra, *Anal. Bioanal. Chem.*, **2004**, 378(4): 1112-1123.

[31] Roussel, C., Dayon, L., Lion, N., Rohner, T. C., Josserand, J., Rossier, J. S., Jensen, H., Girault, H. H., Generation of mass-tags by the inherent electrochemistry of electrospray for protein mass spectrometry, *J. Amer. Soc. Mass Spectrom.*, **2004**, 15(12): 1767-1779.

[32] Dayon, L., Roussel, C., Prudent, M., Lion, N., Girault, H. H., On-line counting of cysteine residues in peptides during electrospray ionisation by electrogenerated tags and its application to protein identification, *Electrophoresis*, **2005**, 26(1): 238-247.

Chapter 7. Evaluation of mass spectrometric front-end technologies to increase peptide uniqueness in shotgun proteomics

1. Introduction

Protein identification by mass spectrometry mostly proceeds through bottom-up strategies, in which protein identification is derived from the analysis of proteolytic peptides, contrary to top-down strategies in which the analysis proceeds from intact proteins. Since top-down strategies are beyond the scope of this work, and necessitate the development of dedicated instrumentation based on FT-ICR-MS, this section focuses on bottom-up approaches (the reader is referred to references [1-7] for more details on the top-down strategy). In bottom-up approaches, proteins are enzymatically digested (most often by trypsin [8]) and the resulting peptides are subjected to mass spectrometric analysis. If the tryptic peptide mixture comes from very simple protein mixtures, their mass alone is sufficient to identify their parent proteins. This is the peptide mass fingerprinting approach [9], which is mostly used with MALDI-TOF-MS after protein separation by two dimensional gel electrophoresis (2D-GE). As summed up in chapter 1, another approach has been introduced recently to circumvent the caveats of protein separations: the whole protein

mixture is digested by trypsin to produce a peptide mixture that is more complex in terms of component number, but more homogeneous in terms of physical properties (hydrophobicity, charge, size…) and can thus be handled more easily. This approach is the shotgun strategy pioneered by John Yates [10-12], and Ruedi Aebersold [13-15] for its quantitative aspects. Unfortunately, the peptide mass fingerprinting approach (identification of a protein from the measurement of its tryptic peptide masses alone) proved to be inefficient when dealing with complex protein mixtures (and thus even more complex tryptic peptide mixtures), and online MS/MS peptide sequencing is necessary to complement the peptide masses information with partial sequence information to interrogate genomic or proteomic databases. This shotgun approach is best implemented with ion trap instruments that allow fast MS/MS sequencing with decent mass accuracy, in the context of discovery proteomic experiments. One caveat of such experimental procedure is that it requires long acquisition times (typically 24 hour run of LC/LC-MS/MS [11] followed by days to weeks of database interrogation). The second drawback of the method is that it leads to complete re-identification of proteins even when the protein content of a sample is known (as is the case when one wants to compare a disease sample to a reference one).

Recently Richard Smith has introduced a new methodology called Accurate Mass Tags approach, in which a proteome is first analysed in a discovery mode by “traditional” shotgun techniques (LC-MS/MS) on an ion trap instrument. The set of identified peptides is then analysed to mine out those peptides that are unique in the digested proteome (i.e. that do not have any isobar) at high mass resolution. Other comparable samples (such as a disease sample) can thus be run on fast LC-FT-ICR-MS (without any MS/MS): only the peptides previously identified as unique are looked for by the mass spectrometer and leads

to protein presence confirmation [16-20]. A simulation of the number of unique peptides based on high mass resolution measurements is shown in Figure 119:

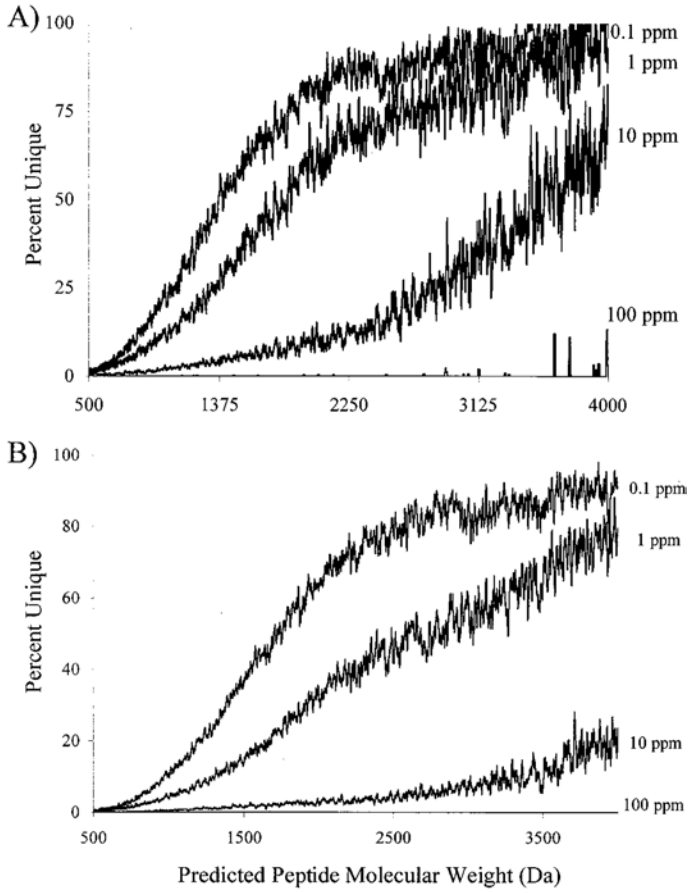


Figure 119. Calculation of unique peptides in A) yeast and B) C. Elegans (from NCBI GenBank database) at various mass accuracies. Reprinted with permission from [16].

Unique peptides are here defined as peptides whose sequence can be unambiguously assigned from mass spectral measurement and database knowledge, be it a comprehensive community database such as SwissProt or Trembl, or an in-house experimental database. The goal of all shotgun approaches is to acquire as much information as possible before database interrogation in order to limit the number of protein candidates that would match the experimental data. Acquiring MS/MS data is one way to limit the search space within the protein database, as well as working with accurate mass tags at high mass accuracy and high mass resolving power allows to provide more stringent information to database search.

High mass accuracy and resolving power, and partial sequence information are not the only ways to add discriminating information to database search, and additional methodologies are being examined in order to complement mass spectrometric information.

2. Complementary technologies in shotgun proteomics

Two main methodologies have been investigated to complement mass spectrometric information: peptide retention time in reverse phase liquid chromatography, and isoelectric point measurement.

2.1. Peptide retention times

Richard Smith's group has attempted to complement the AMT strategy based on high resolution mass spectrometry with peptide retention times in reverse phase liquid chromatography. In order to evaluate the potential of the approach, Petritis *et al* have developed a model for the prediction of peptide retention times based on a neural network trained on a set of 42378 tryptic peptides from *S. Oneidensis* [21]:

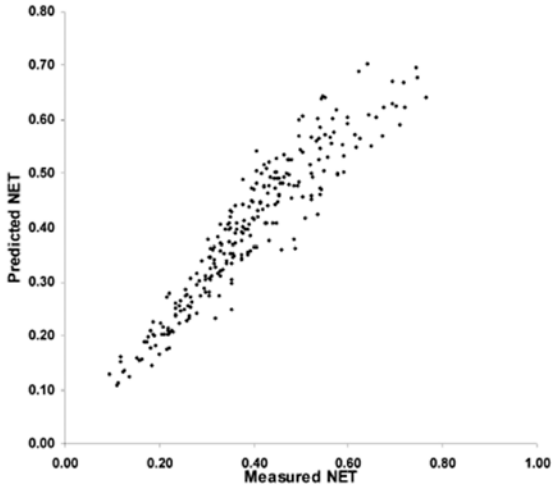


Figure 120. Predicted Net Retention Time versus observed Net Retention Time. Reprinted with permission from [21].

Authors then used the predicted net retention time to re-analyse the number of accurate mass tags based on high mass accuracy previously measured in *D. Radiodurans*:

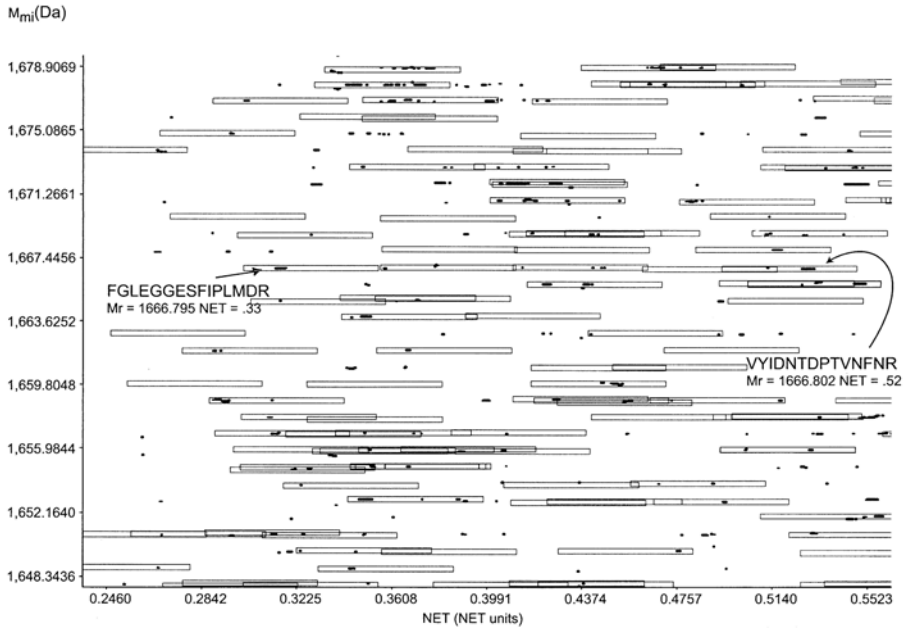


Figure 121. Small region of a 2D map of a TOF-MS analysis of D. Radiodurans. Each spot corresponds to an observed peptide; boxes represent the expected elution times (with tolerance derived from experimental variance). Reprinted with permission from [22].

The addition of LC retention times allows to discriminate between peptides that have nominal same masses at a given resolution (spots spanned on the same horizontal line): for example peptides FGLEGGESFIPLMDR and VYIDNTDPTVNFNR have masses differing by 4.2 ppm only, but differ in retention time and are thus clearly resolvable. The main drawback of this approach is that peptide retention times are not truly orthogonal to peptide masses, as shown below:

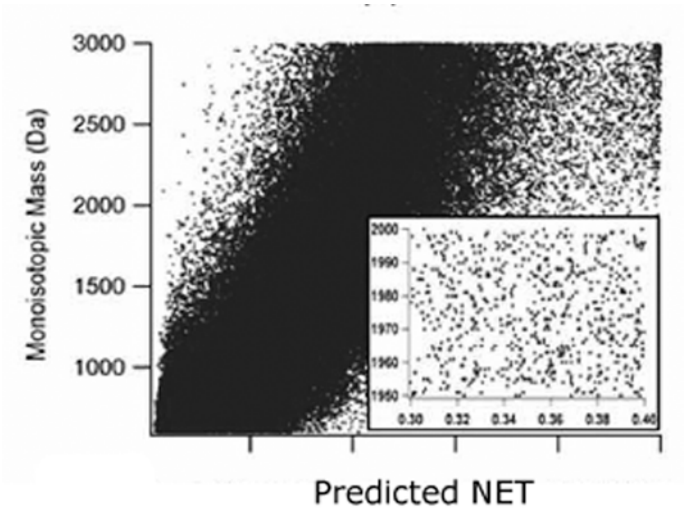


Figure 122. Predicted molecular weight and Net Retention Times of tryptic peptides from D. Radiodurans. Reprinted with permission from [23].

The use of peptide retention times to increase the number of unique peptides in whole digested proteomes is shown in Figure 123:

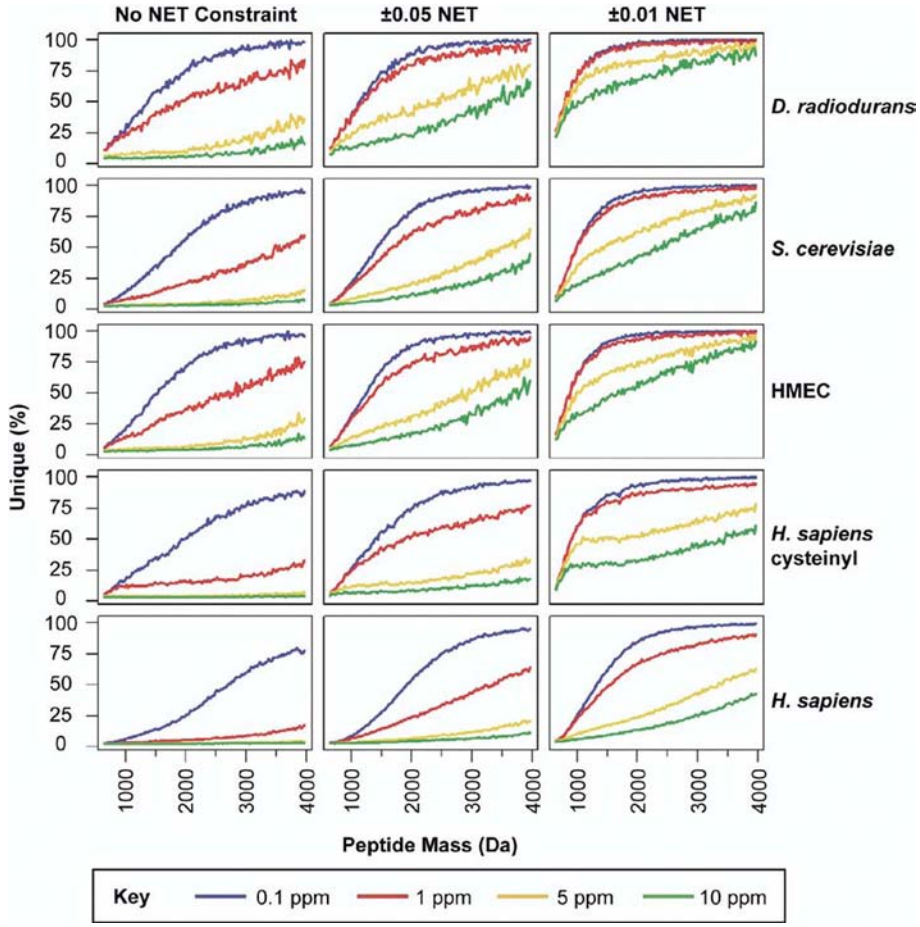


Figure 123. Percentage of unique peptides in whole proteomes based on mass accuracy alone (left) and different peptide retention time stringency (middle and right). HMEC is a Human Mammary Epithelial Cells proteome. Reprinted with permission from [23].

The addition of the retention time information allows to increase the percentage of unique peptides in each proteome. However, as the methodology does not lead to a dramatic increase in the number of unique peptides, authors conclude that this methodology can prove efficient for low complexity proteomes or for sub-proteomes only, and that the measurement of higher organisms' proteomes by this methodology is still out of reach.

2.2. Isoelectric point

Similarly to peptide retention times, Stephenson's group has investigated the use of isoelectric point to increase the number of unique peptides in shotgun proteomics [24].

Tryptic peptide molecular weight and isoelectric points were computed and the number of unique peptides was counted:

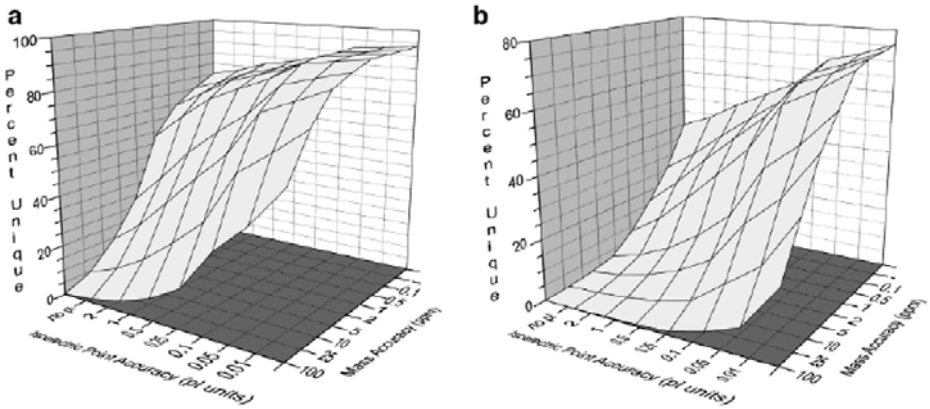


Figure 124. Percentage of unique peptides versus mass and isoelectric point accuracies for E. Coli (a) and R. Norvegicus (b). Reprinted with permission from [24].

Mass accuracies are varied between 0.1 and 100 ppm and isoelectric point accuracies between 0.01 and 2. Results are similar to those exposed below, and were further exploited experimentally with in-gel isoelectric focusing [25]. The limit of the approach was the actual resolution of the IEF, which was unknown. In practice, 28 fractions were cut over the IPG strip (pH 3-10), corresponding to a theoretical ΔpH= 0.25. Authors failed to identify one source of error that was misinterpreted as diffusion of peptides within the gel once the voltage was switched off: in fact some peptides have a charge derivative at isoelectric point close to zero, which means that they will focus very slowly compared to other peptides. Hence, they will produce broad bands unless infinite focusing time is applied [26] (see below for a further discussion on the limits of isoelectric point measurements).

3. Online electrochemical tagging of cysteine residues: eTag(Cys)

Our laboratory has recently introduced an online tagging technology [27] that allows counting of cysteines contained in a peptide from a simple MS spectrum [28]. Basically, neutral hydroquinones are injected together with the sample to analyse. Due to their low electrochemical potential, hydroquinones are converted into benzoquinones on the electrospray electrodes, and reacts specifically with cysteine residues:

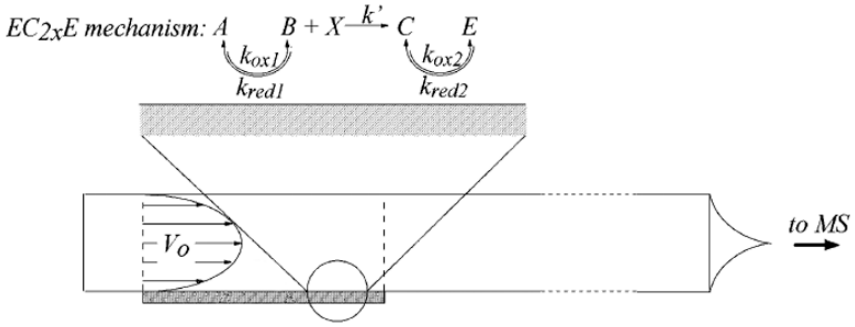


Figure 125. Scheme of the electrochemical tagging: the hydroquinone A is converted into benzoquinone B onto the electrospray electrode, then reacts with the cysteine containing peptide X to give the adduct C that can be further reoxidised into E. Reprinted with permission from [29].

The overall reaction is an ECE (electrochemical-chemical-electrochemical) mechanism that has been studied in details [30]:

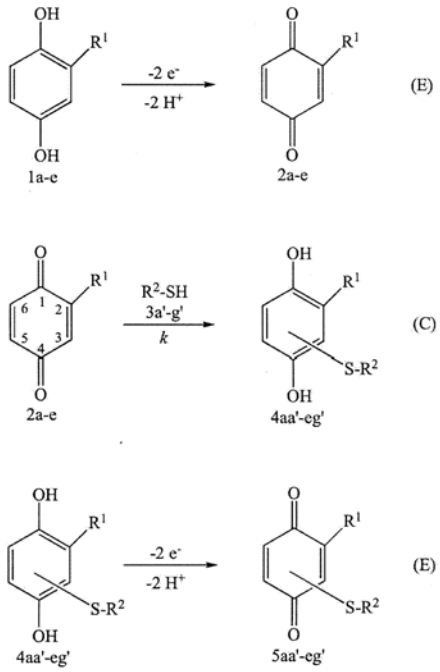


Figure 126. Mechanism of the ECE reaction. Reprinted with permission from [30].

The addition of benzoquinone onto the cysteine residue proceeds through a Michael addition that is specific in acidic conditions. Both the design of the microspray emitter (whose geometry plays a role on the overall tagging efficiency) [29] and the choice of the hydroquinone probe [31] have been optimised to reach good tagging conditions, in which the tagging reaction is not quantitative, but is efficient enough to tag several cysteine residues in the same peptide. The typical tagging result is depicted in Figure 127:

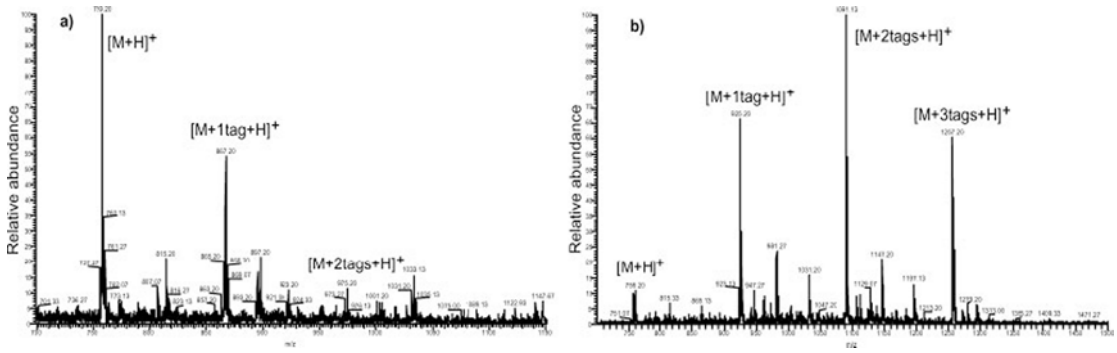


Figure 127. Tagging of the ACKCTCM peptides by hydroquinone (a) and carboxymethylhydroquinone (b). Reprinted with permission from [28].

The tagging process proved efficient even with consecutive cysteines [28], and allows direct counting of cysteines in a peptide by examining the number of additional peaks in the mass spectrum [32]. When used for database interrogation for protein identification by peptide mass fingerprinting, this additional information was shown to result in a dramatic increase in protein identification scores [28].

4. Off-Gel electrophoresis of peptides and proteins

Our laboratory has introduced a new technology for liquid-phase separation of peptides and proteins based on isoelectric focusing [33, 34]. Basically, wells are put over an immobilised pH gradient (IPG) strip, in which the solution to fractionate is sampled. Upon application of an electric field, peptides or proteins moves within wells according to their charge until they reach their isoelectric point, and enter the IPG gel only to jump from one well to another.

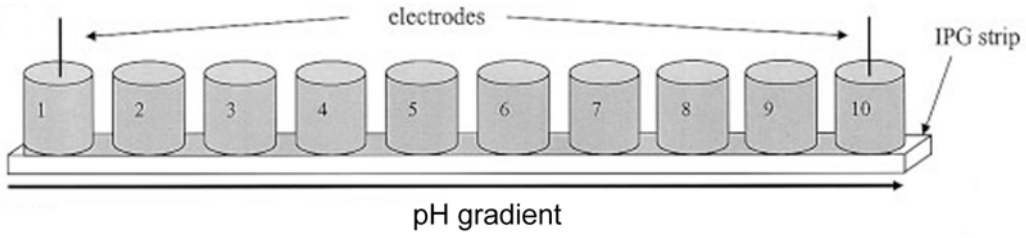


Figure 128. Scheme of the Off-Gel electrophoresis device. Reprinted with permission from [34].

The clear advantage of this separation procedure over in-gel separations is that peptides or proteins can be directly recovered from solution. Its advantages over capillary isoelectric focusing are a higher loading capacity and an easier fraction recovery. Recently, this technique has been used for human plasma proteome fractionation and subsequent tryptic peptide separation prior to LC-MS/MS [35]. In this work, the peptide p*I* was used to validate MS/MS spectra by comparing the p*I* predicted from the sequence obtained by

MS/MS with the experimental one. The resolution of Off-Gel electrophoresis depends on the IPG pH gradient and on the number of wells placed over the IPG strip. Depending on the IPG strip used and the focusing time applied, resolutions of 0.2 can be reached in routine and 0.05 in particular experiments.

5. Computing of peptide uniqueness

In order to evaluate the potential interest of online cysteine tagging (eTag(Cys)) and isoelectric separation (Off-Gel) for accurate mass tag approaches, a program was written using Igor software to simulate whole proteome digestion, compute peptide molecular weights and p*Is*, and count the number of unique peptides within the whole digested proteome according to the mass accuracy used, p*I* accuracy, and number of cysteines.

5.1. Material and methods

The program was written using Igor (WaveMetrics, Portland, USA) and run on an Apple PowerBook G4 (1.33 GHz and 512 Mo SDRAM) or on an Apple iMac G5 (2 GHz and 1 Go SDRAM). Proteomes of *Deinococcus Radiodurans*, *Saccharomyces Cerevisiae*, and *Homo Sapiens* were downloaded from SwissProt in the fasta format. The program can simulate protein digestion by trypsin (cleavage at lysine or arginine) or by pepsin (cleavage probability of 0.8 at F, L, and E and user-defined probability at other residues). Other digestion parameters, such as the number of allowed tryptic miscleavages, the minimal peptide length, the pepsin cleavage frequency… are user-defined and can be adjusted for each simulation. Proteins and peptides molecular weight and p*I* are computed from the following values:

Amino acid	COOH p*K*a	NH3 p*K*a	Side chain *p*Ka	Monoisotopic molecular weight (Da)	Average molecular weight (Da)
Alanine (A)	2.4	9.9		71.03711	71.0788
Arginine (R)	1.8	9.0	12.2	156.10111	156.1875
Asparagine (N)	2.0	8.8		114.04293	114.1038
Aspartic acid (D)	2.0	10.0	3.9	115.02694	115.0886
Cysteine (C)	1.9	10.3	8.4	103.00919	103.1388
Glutamine (Q)	2.2	9.1		128.05858	128.1307
Glutamic acid (E)	2.1	10.0	4.3	129.04259	129.1155
Glycine (G)	2.4	9.8		57.02146	57.0519
Histidine (H)	1.8	9.2	6.1	137.05891	137.1411
Leucine/Isoleucine (I/L)	2.3	9.7		113.08406	113.1594
Lysine (K)	2.2	9.2	10.8	128.09496	128.1741
Methionine (M)	2.2	9.3		131.04049	131.1926
Phenylalanine (F)	2.6	9.2		147.06841	147.1766
Proline (P)	2.0	10.6		97.05276	97.1167
Serine (S)	2.2	9.4		87.03203	87.0782
Threonine (T)	2.1	9.1		101.04768	101.1051
Tryptophane (W)	2.4	9.4		186.07931	186.2132
Tyrosine (Y)	2.2	9.1	10.1	163.06333	163.1760
Valine (V)	2.3	9.7		99.06841	99.1326

Table 18. Physicochemical parameters of amino acids. Monoisotopic mass of a water molecule is 18.01056 Da, average mass of a water molecule is 18.01524.

5.2. Program validation

The program was validated manually on a small set of proteins indicated below against the online tool MS Digest (http://prospector.ucsf.edu):

Sequence Or SwissProt reference	Number of tryptic peptides from MS digest	Number of tryptic peptides from the in-house software
E(31)	1	1
P39687	65	65
P38903	216	221
P31383	162	162
Q00362	151	159

Table 19. Validation set of the in-house software.

Differences in the number of tryptic peptides produced by the two programs correspond to the fact that trypsin cleavage inhibition by proline is not taken into account in

the in-house program. From the validation set, it is estimated that this results in an additional 5% of tryptic peptides. All other tryptic peptides are identical. Monoisotopic and average molecular weights were found to be equal to those calculated by MS Digest. p*I*s were validated against the online tool p*I*/MW Compute (http://www.expasy.ch): due to differences in p*K*as used (in particular the influence of amino acids adjacent to the N-ter or C-ter on N-ter or C-ter p*K*as), the p*I* values were slightly different, but never differed by more than 0.5 p*H* unit.

5.3. Note on mass accuracy and mass resolution

Mass accuracies are expressed below in part per million (ppm) and are defined as:

$$MA = \frac{\left| m_{measured} - m_{theoretical} \right|}{m_{theoretical}} \times 10^6$$

Mass accuracy solely expresses the trueness of mass measurements, but does not relate to resolving power (peak sharpness), which is defined as:

$$MRP = \frac{m_{measured}}{FWHH}$$

where FWHH stands for full width at half height. In order to give an idea of the separation power of mass analysers, a peak capacity (PC) can be defined as:

$$PC = \frac{m_{\max} - m_{\min}}{FWHH}$$

State-of-the-art mass accuracies and mass resolving power and peak capacities of various instruments are given in Table 20:

	Mass accuracy (ppm)	Mass resolving power	Peak capacity
Ion trap	100-200	2000	20000-40000
TOF	5-50	>10000	∞
Orbitrap	<10 ppm	>50000	>200000
FT-ICR	1-10 ppm	>100000	>400000

Table 20. State-of-the-art performances of various analysers. Ion trap and TOF data are from our lab, Orbitrap data is from [36] and FT-ICR from chapter 5.

All simulations performed below are based on mass accuracy alone; effects due to limitations of peak capacities are not taken into account.

5.4. Whole protein analysis

The three proteomes tested come from different taxonomic species (bacteria, yeast and primates):

	D. Radiodurans	S. Cerevisiae	H. sapiens
Number of proteins	3084	6229	12860

Table 21. Number of proteins for each organism.

Protein molecular weights are distributed over a wide range, as shown in Figure 129.

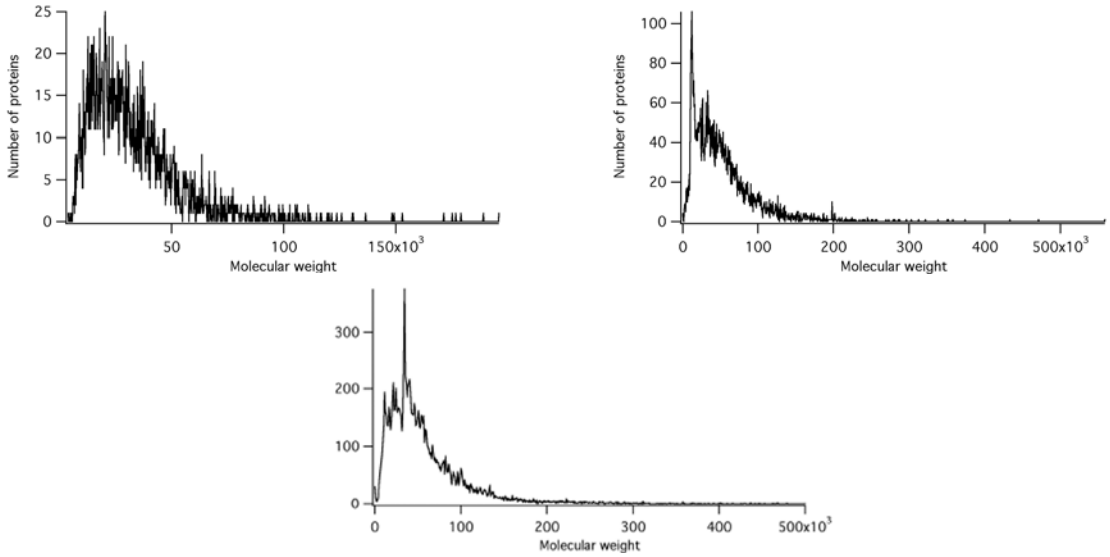


Figure 129. Protein molecular weight histograms for Deinococcus Radiodurans (top left), Saccharomyces Cerevisiae (top right) and Homo sapiens (bottom).

Protein isoelectric point distributions display the well-known bimodal distribution, as shown below:

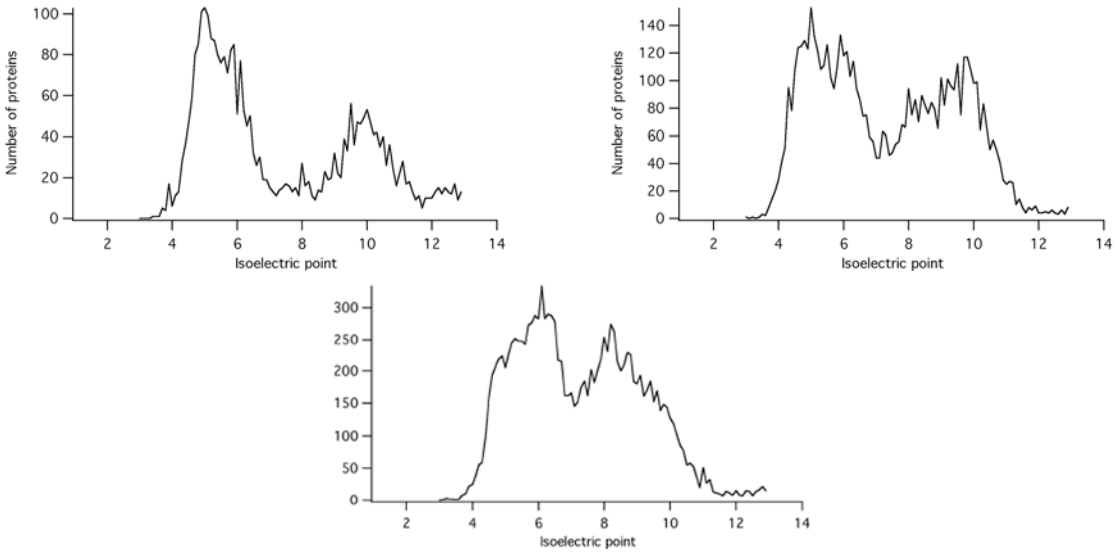


Figure 130. Protein isoelectric point histograms for Deinococcus Radiodurans (top left), Saccharomyces Cerevisiae (top right) and Homo sapiens (bottom).

These bimodal distributions were primarily thought to be the result of evolution to intracellular pHs. Interestingly, Weiller *et al* have recently shown that this bimodal distributions are a result of amino acids p*K*as only [37].

5.5. Simulation of trypsin digestions

The three above-mentioned proteomes were digested by trypsin, with two miscleavage sites per peptide allowed and a minimal peptide length of four residues.

	D. Radiodurans	S. Cerevisiae	H. Sapiens
Number of tryptic peptides	254133	884737	2047516
Average number of tryptic peptides per protein	82.4	142.0	159.2

Table 22. Results of whole proteome tryptic digestions.

Tryptic peptides isoelectric point span over the whole pH range, as shown below:

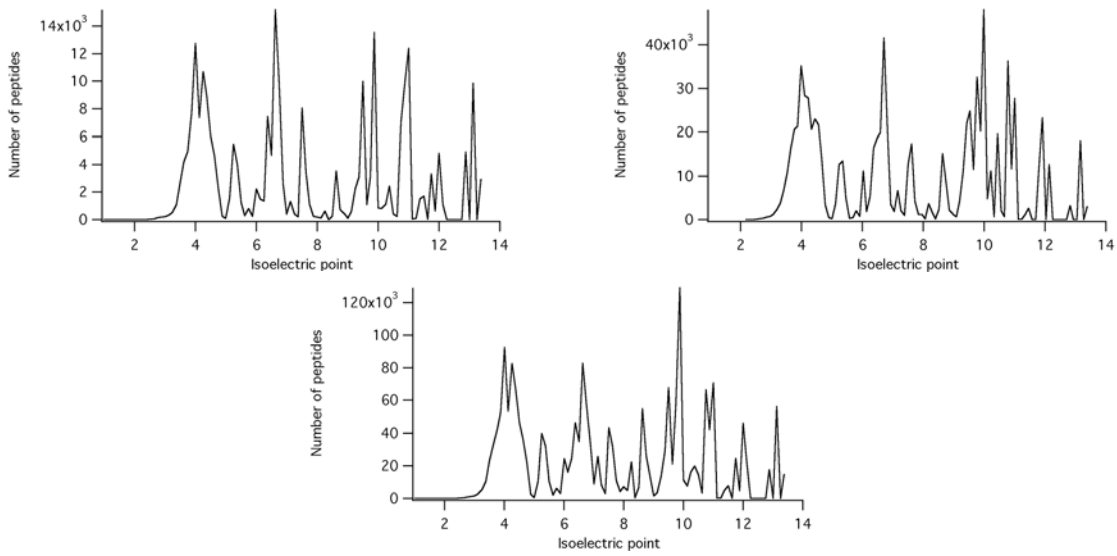
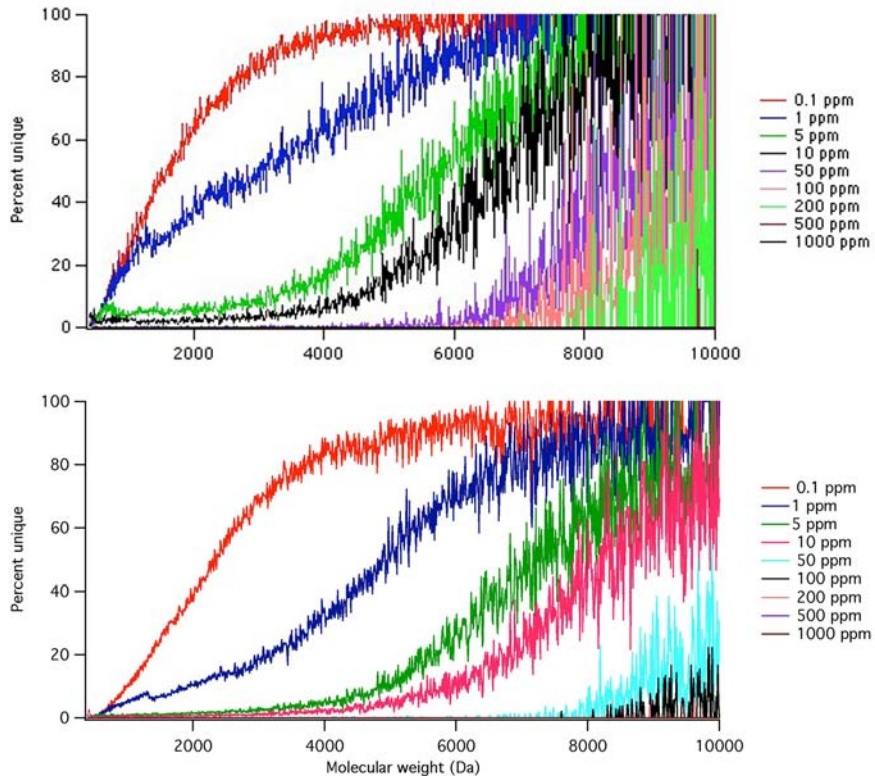


Figure 131. Isoelectric points of tryptic peptides from whole proteomes for D. Radiodurans (top left), S. Cerevisiae (top right) and H. Sapiens (bottom).

5.6. Peptide uniqueness based on mass accuracy alone

The number of unique peptides in each whole proteome was computed based on mass accuracy alone. Figure 132 shows the percentage of unique peptides for each zone of 10 Da at various mass accuracies.



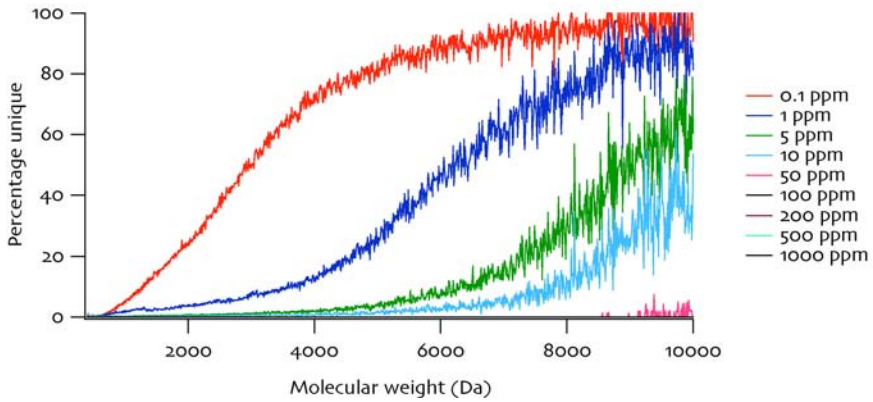


Figure 132. Percentage of unique peptides for each 10 Da bin for D. Radiodurans (top), S. Cerevisiae (middle) and H. Sapiens (bottom).

The noisy parts of the curves at high mass are due to the fact that there are few peptides in each 10 Da bin. As expected, the higher the mass resolution, the more unique peptides. Figure 133 present the average number of unique peptides per protein for each proteome.

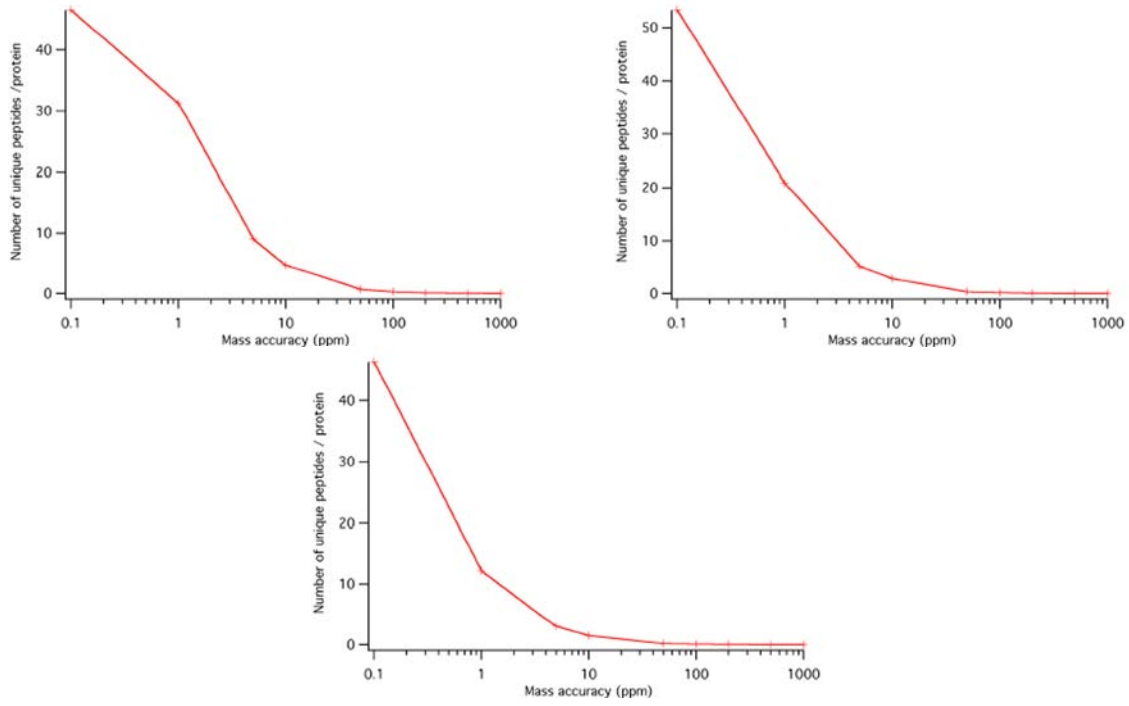


Figure 133. Average number of unique peptides per protein based on mass resolution for D. Radiodurans (top left), S. Cerevisiae (top right) and H. Sapiens (bottom).

As can be seen in Figure 133, even at 10 ppm mass accuracy which is achievable only on best instruments in routine (see Table 20), no more than approximately 5 unique peptides per protein can be found. And as the mass accuracy decreases (above 10 ppm), the average number of unique peptide per protein rapidly falls down to zero.

5.7. Effect of cysteine count on peptide uniqueness

Digested proteomes contain a large number of peptides free of cysteines (225327 in D. Radiodurans, i.e. 88.6 % of tryptic peptides, 719126 in S. Cerevisiae, i.e. 81.3% of tryptic peptides, and 1454607 in H. Sapiens, i.e. 71.0% of tryptic peptides). This low number of cysteine containing tryptic peptides is at the core of the Isotope Coded Affinity Tags (ICAT) approach developed by Ruedi Aebersold to simplify tryptic peptides complex mixtures, since selecting only those peptides leads to a 5 to 10 fold simplification of the mixture [13-15]. Figure 134 shows the orthogonality between the number of cysteines per peptide and the peptide molecular weights. Surprisingly, it does not appear to be any correlation between the number of cysteines and the peptide molecular weight.

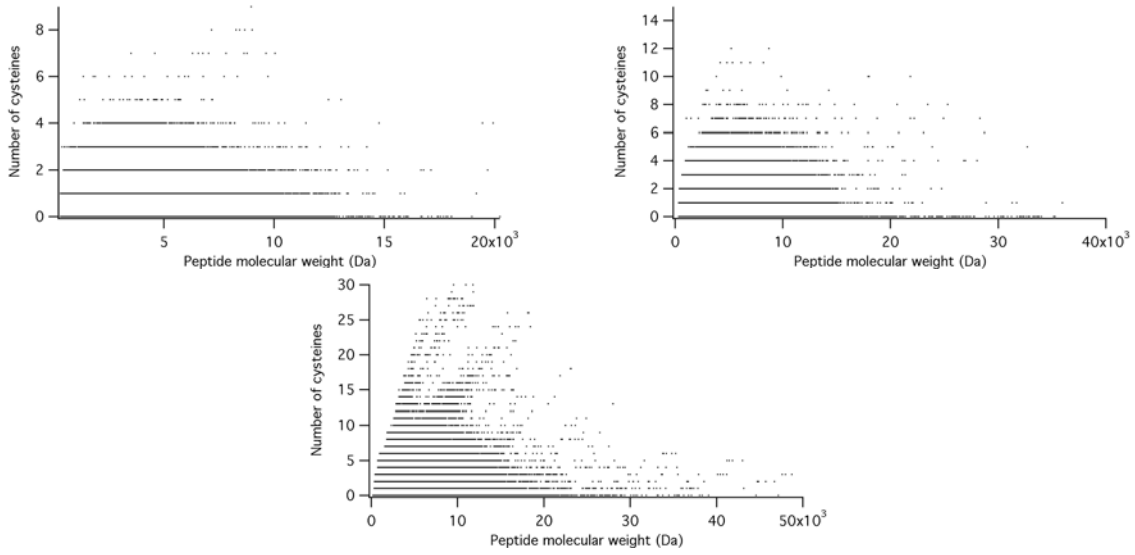


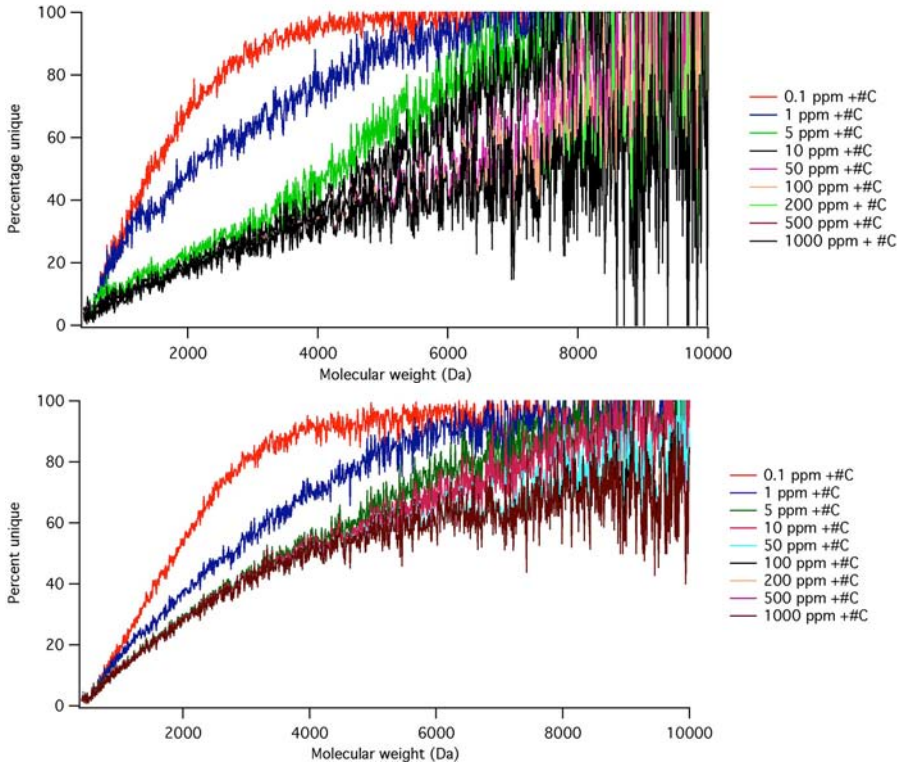
Figure 134. Orthogonality of molecular weight and number of cysteines information for D. Radiodurans (a), S. Cerevisiae (b) and H. Sapiens (c).

However, the maximum number of cysteines contained in a single tryptic peptide exceeds by far the potential of the online electrochemical tagging presented above:

	D. Radiodurans	S. Cerevisiae	H. Sapiens
Max number of cysteine per tryptic peptide	9	22	56
Average number of cysteines per peptide	0.14 ± 0.44	0.25 ± 0.61	0.46 ± 0.99

Table 23. Statistics of cysteine content in tryptic peptides.

However the simulation was performed as if the technology was able to count the number of cysteines without any limitation. The percentage of unique peptides per bin of 10 Da is presented in Figure 135 when the number of cysteines is taken into account.



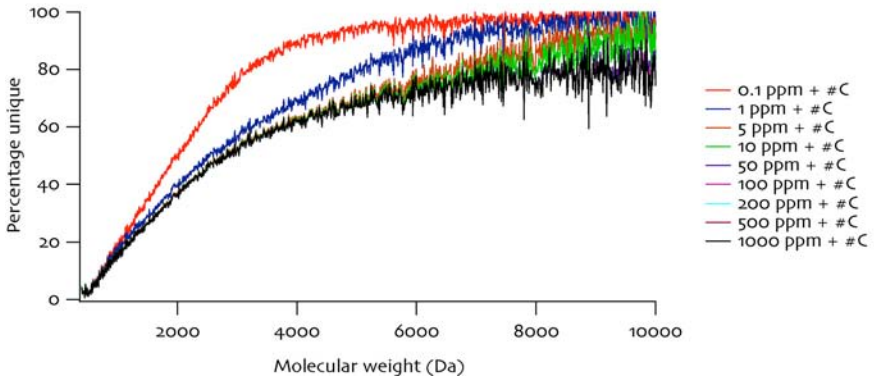


Figure 135. Percentage of unique peptides for each 10 Da bin for D. Radiodurans (top), S. Cerevisiae (middle) and H. Sapiens (bottom).

For ultrahigh mass accuracy (0.1 ppm that may be reachable only on FT-ICR, see Table 20), the information about the number of cysteines does not bring much; but for lower accuracies, the number of cysteines increases the number of unique peptides and tends to level the effect of mass accuracy: there is indeed not much difference in H. Sapiens between the percentage of unique peptides at 200 ppm (typical of an ion trap instrument) and 5 ppm (more in the range of a good TOF instrument). Figure 136 shows the average number of unique peptides per protein for each proteome:

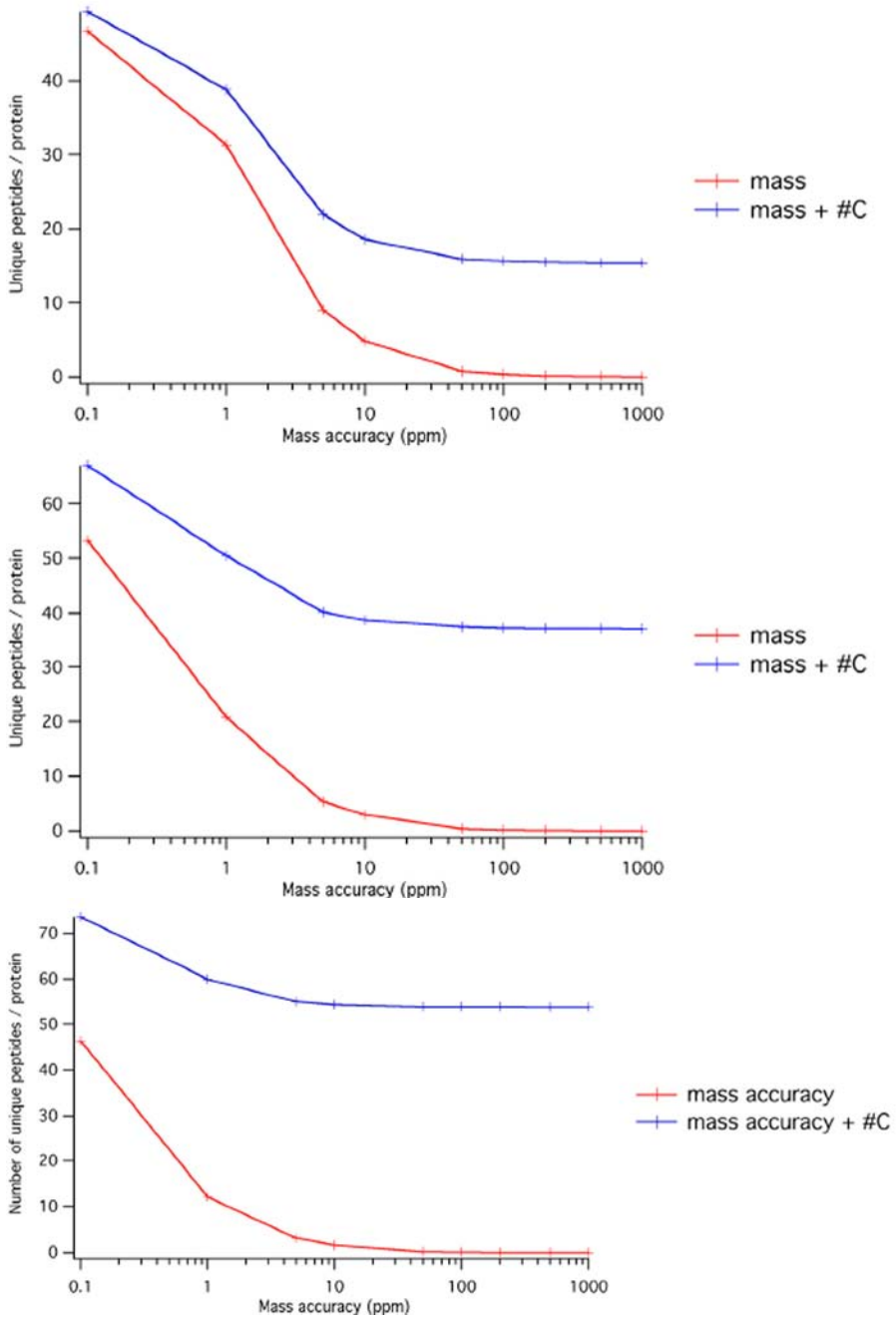


Figure 136. Average number of unique peptides per protein based on mass resolution and cysteine number per peptide for D. Radiodurans (top), S. Cerevisiae (middle) and H. Sapiens (bottom).

Interestingly, at high mass accuracy, there is no clear increase in the number of unique peptides per protein compared to using mass accuracy alone for *D. Radiodurans*, whereas for *S. Cerevisiae* and *H. Sapiens* there is a clear increase in unique peptides even at very high mass accuracy (see Figure 133 for comparison). At low mass accuracy, the

addition of the amount of cysteine in each peptide dramatically increases the number of unique peptides; not only is it improved, it plateaus at approximately 18 peptides per protein for *D. Radiodurans*, 40 peptides per protein for *S. Cerevisiae*, and 55 peptides per protein in *H. Sapiens*.

Given the results presented above, the eTag(Cys) methodology has the potential to greatly improve peptide identification and is especially efficient when combined with relatively low mass accuracy mass analysers (such as ion traps). The great advantage of the method is that it does not add any additional experiment: it only requires the addition of carboxymethylhydroquinone to the solution delivered to the mass spectrometer.

5.8. Effect of pI on peptide uniqueness

Isoelectric point of proteins is known to be orthogonal to mass for proteins (which is a great strength of two-dimensional gel electrophoresis), but there is not much documentation at the peptide level.

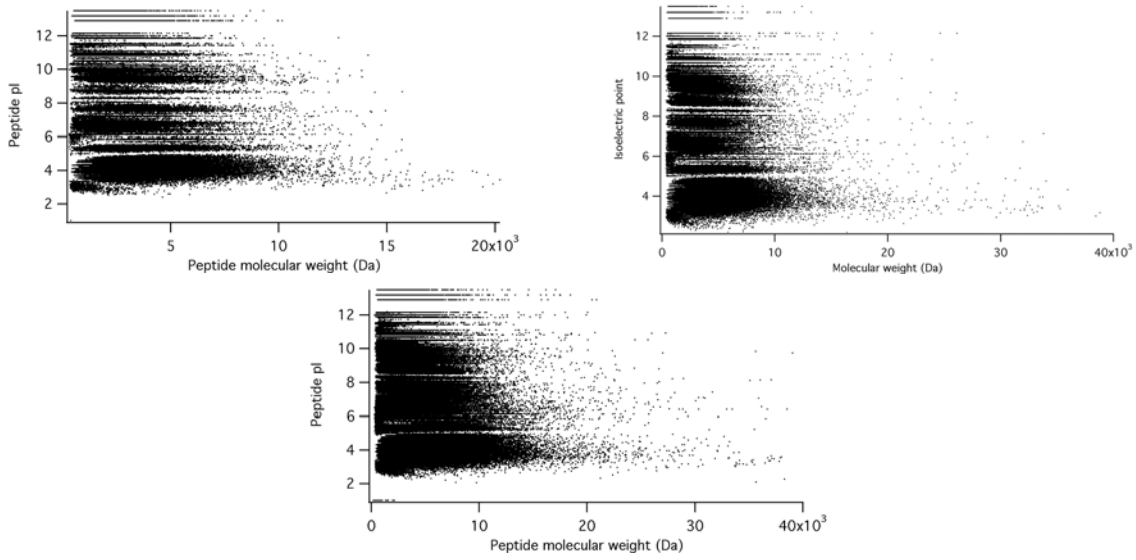


Figure 137. Distribution of peptide isoelectric point versus peptide molecular weight for D. Radiodurans (a), S. Cerevisae (b) and H. Sapiens (c).

Contrary to peptide retention times in liquid chromatography (see Figure 122), p*I*s are independent of molecular weights. The percentage of unique peptides for each 10 Da

bin is presented in Figure 138, based on mass accuracy and peptide isoelectric point (for a p*I* accuracy of 0.5 pH unit) .

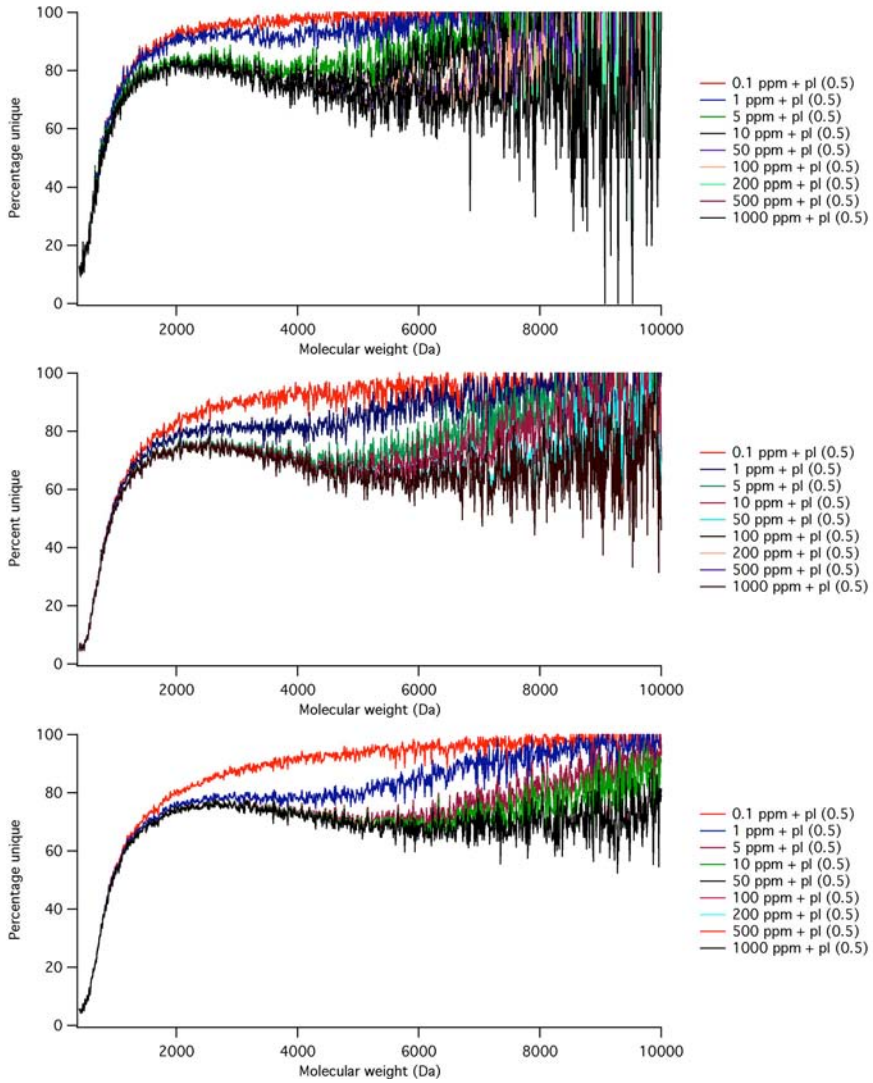


Figure 138. Percentage of unique peptides for each 10 Da bin for D. Radiodurans (top), S. Cerevisiae (middle) and H. Sapiens (bottom) based on mass accuracy and peptide pI.

As can be seen in Figure 138, the addition of isoelectric point for the discrimination of isobaric peptides has a dramatic effect on the percentage of unique peptides (see Figure 132 for comparison). At all mass accuracies the percentage of unique peptides is increased and is in the range 60-80% above 2000 Da in all digested proteomes. The translation of this increase in terms of average number of unique peptides per protein is shown in Figure 139:

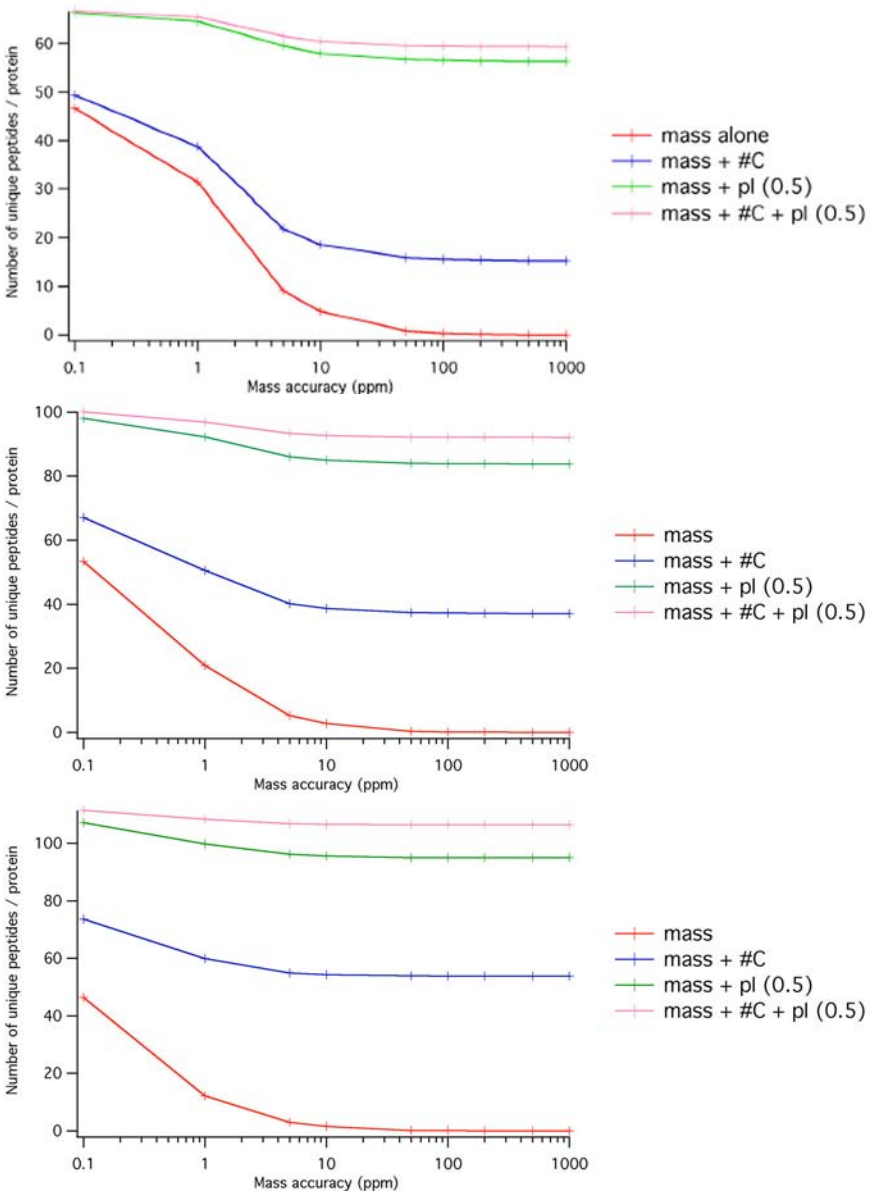


Figure 139. Average number of unique peptides per protein, taking into account mass accuracy, number of cysteine per peptide and isoelectric point for D. Radiodurans (top), S. Cerevisiae (middle) and H. Sapiens (bottom).

As clearly visible from Figure 139, the addition of isoelectric point to discriminate peptides greatly increases the average number of unique peptides per protein: even at low mass accuracy, there are approximately 55 unique peptides per protein for *D. Radiodurans*, 85 for *S. Cerevisiae* and 95 for *H. Sapiens*. Moreover, the addition of the number of cysteines as a discriminating parameter further increases the number of unique peptides. The p*I* accuracy of 0.5 pH unit can be reached in routine with Off-Gel electrophoresis or

other techniques, and one can push to better resolutions. Figure 140 shows the influence of isoelectric point accuracy on the percentage of unique peptides:

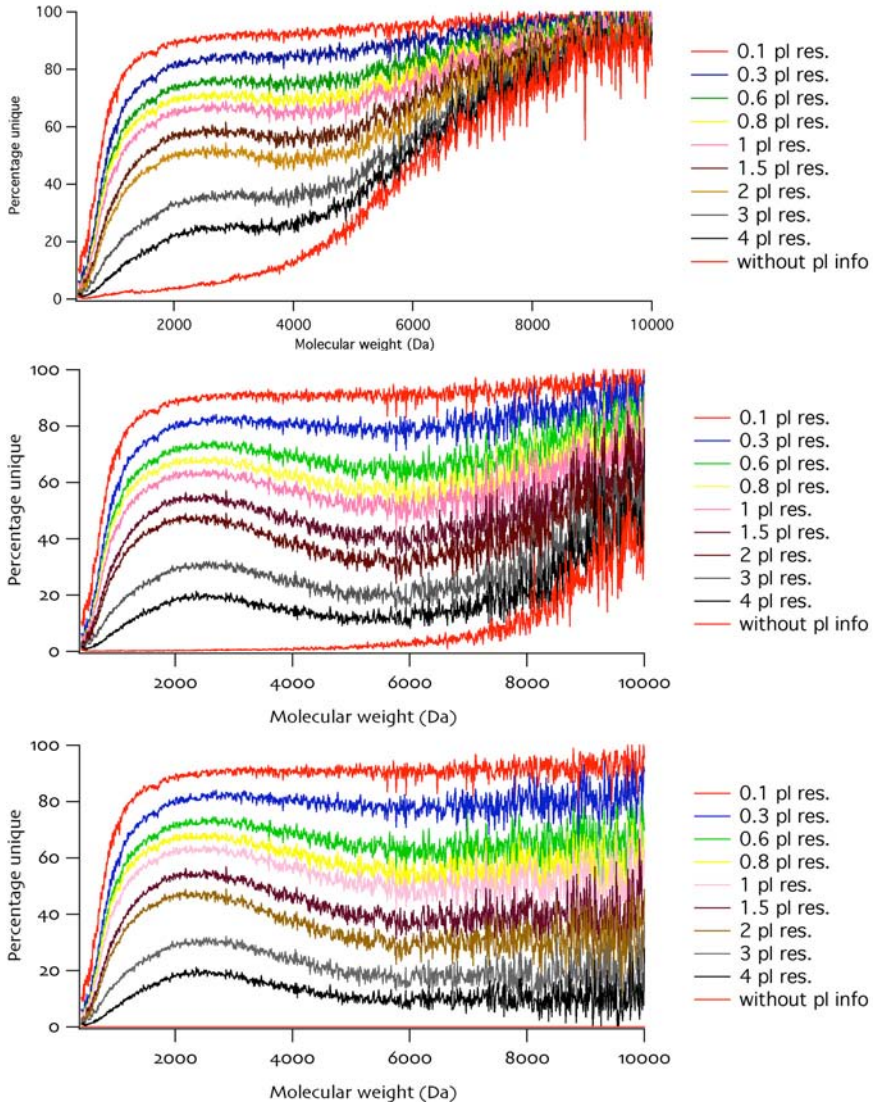


Figure 140. Percentage of unique peptides for each 10 Da bin in H. Sapiens digested proteome for various pI resolution at different mass accuracies: 1 ppm (top), 10 ppm (middle) and 100 ppm (bottom).

The first important result is that even when working with high mass accuracy, adding the isoelectric point information increases the percentage of unique peptides, as shown in the top row of Figure 140. The second result is that adding isoelectric point information to low mass accuracy measurements (e.g. 100 ppm, which is a good mass accuracy for an ion trap analyser) brings the percentage of unique peptides higher than with

ultrahigh mass accuracy alone (comparison of the 0.1 p*I* resolution curve in the bottom row of Figure 140 to the "without p*I* info" curve in the top row of Figure 140). Lastly, even poor information (low p*I* accuracy) has a strong effect on the number of unique peptides.

6. Limitations of the simulation

The simulation presented above calculate the number of unique peptides in whole digested proteomes downloaded from SwissProt, which means that post-translational modifications are not taken into account. Introducing possible post-translational modifications would increase the total number of peptides, and thus statistically decrease the number of unique peptides. Furthermore, unique peptides identified above may not be readily observable due to poor ionisation efficiency or ion suppression phenomena. Numbers of unique peptides presented above are thus a maximal number of theoretical unique peptides and it is thus of utmost importance to maximise these numbers in order to cope with experimental complexity and limitations.

7. Meaning of peptide uniqueness and the protein inference problem

All the work presented above deals with the identification of unique peptides within digested proteomes based on different measurements (be it peptide masses alone at various mass accuracies, the number of cysteines, peptide isoelectric point…). Unique peptides are peptides whose sequence can be unambiguously assigned given experimental measurements (mass, p*I*, LC retention time, number of cysteines…) within a proteome database (be it

SwissProt, Trembl or an experimental database)[5]. Unique peptides can be compared as peptides sequenced by MS/MS in a more traditional shotgun approach, whose sequence is known and can be used for protein identification. But identifying peptides is only the first step in protein identification, and matching experimental data at the peptide level to protein candidates is a difficult task, known as the protein inference problem. There is no fundamental difference in inferring protein identification from unique peptides (or accurate mass tags) or from partial sequence information obtained by MS/MS. The limitations of protein inference have been reviewed recently by Nesvizhskii and Aebersold [38].

First, different proteins can share a high degree of homology, as for splice variants; for example, human F actin capping protein beta subunit has two known splice variants that share 99.6% of homology, as shown in Figure 141:

```
UserSeq1,      1 SDQQLDCALDLMRRLPPQQIEKNLSDLIDLVPSLCEDLLSSVDQPLKIARDKVVGKDYLL
UserSeq2,      1 SDQQLDCALDLMRRLPPQQIEKNLSDLIDLVPSLCEDLLSSVDQPLKIARDKVVGKDYLL
                 ************************************************************

UserSeq1,     61 CDYNRDGDSYRSPWSNKYDPPLEDGAMPSARLRKLEVEANNAFDQYRDLYFEGGVSSVYL
UserSeq2,     61 CDYNRDGDSYRSPWSNKYDPPLEDGAMPSARLRKLEVEANNAFDQYRDLYFEGGVSSVYL
                 ************************************************************

UserSeq1,    121 WDLDHGFAGVILIKKAGDGSKKIKGCWDSIHVVEVQEKSSGRTAHYKLTSTVMLWLQTNK
UserSeq2,    121 WDLDHGFAGVILIKKAGDGSKKIKGCWDSIHVVEVQEKSSGRTAHYKLTSTVMLWLQTNK
                 ************************************************************

UserSeq1,    181 SGSGTMNLGGSLTRQMEKDETVSDCSPHIANIGRLVEDMENKIRSTLNEIYFGKTKDIVN
UserSeq2,    181 SGSGTMNLGGSLTRQMEKDETVSDCSPHIANIGRLVEDMENKIRSTLNEIYFGKTKDIVN
                 ************************************************************

UserSeq1,    241 GLRSIDAIPDNQKFKQLQRELSQVLTQRQ
UserSeq2,    241 GLRSVQTFADKSKQEALKNDLVEALKRKQ
                 ****     *  *   *   *   *   *
```

Figure 141. Sequence homology of the two F actin human variants.

[5] Please note that with the chosen algorithm, similar peptidic sequences coming from different SwissProt entries (known as shared or degenerate peptides) are not counted as unique peptides.

They differ only in the terminal sequence IDAIPDNQKFKQLQRELSQVLTQRQ (variant 1) and VQTFADKSKQEALKNDLVEALKRKQ (variant 2) and the only way to discriminate between the two variants is to measure a tryptic peptide coming from one of these two sequences. Unfortunately, none of the possible tryptic peptides coming from these two sequences is a unique peptide in our simulations, that is to say that the unique peptide approach fails to discriminate between these two protein splice variants.

A more challenging situation is when splice forms originate from skipped exons at the 3' or 5' end: discrimination between different splice forms thus relies on the absence of a given peptide in the experimental dataset, which is always questionable. In practise it is impossible to discriminate between such variants without information about the intact protein (molecular weight or isoelectric point for example).

An even more complex problem is when a whole protein family shares a high degree of homology. Figure 142 shows an example taken from reference [38], in which 11 peptides belonging to proteins from the alpha-tubulin family were identified, but could not be assigned to a particular protein:

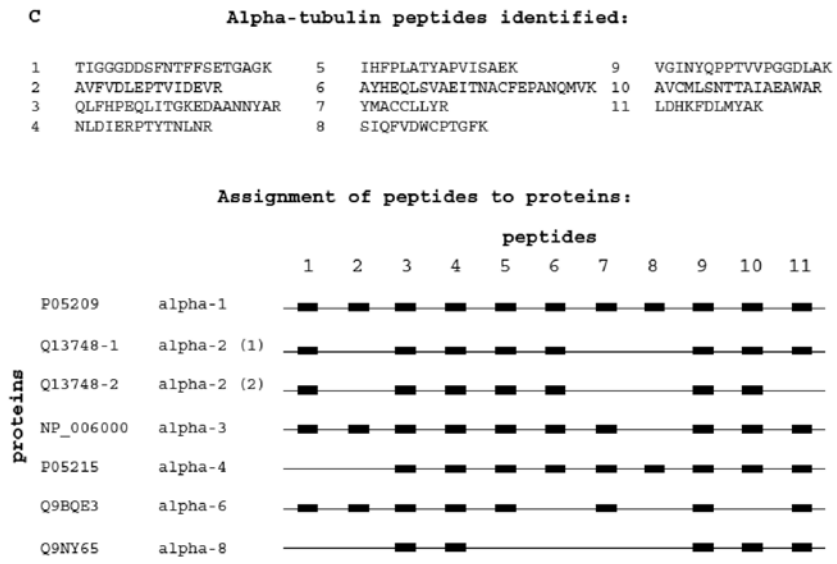


Figure 142. Assignment of measured peptides to proteins belonging to the alpha-tubulin family. Reprinted with permission from [38].

In fact given the experimentally identified peptides, it is impossible to identify which members of the family are present in the sample.

Another challenge in assembling large scale peptide information into protein identifications is the presence of redundancy in databases (same proteins listed under different names or accession numbers) as well as sequence variations that can be due whether to DNA sequencing errors or to real single nucleotide polymorphisms.

The simulations reported in this chapter show that acquiring more physico-chemical information at the peptide level (number of cysteines, isoelectric point, retention time in LC…) can dramatically improve the number of unique peptides per protein (see Figure 139), and thus the sequence coverage, which is a key criteria to increase the probability of discrimination between variants. For the sake of comparison, it is estimated that 30% of identified proteins in a traditional shotgun experiment (based on MS/MS sequencing of peptides) are identified by one peptide only [38]. Classical shotgun approaches as well as accurate mass tags or unique peptides approaches consists more in a gene-product

identification than in true protein identification [39]. In order to convert them into true protein identification technologies, it is of utmost importance to increase the number of peptides identified per protein; Figure 139 shows that combining eTag(Cys) and Off-Gel technologies with 100 ppm mass accuracy measurements has the potential to switch from less than five identified peptides per protein (mass measurement alone) to 100 peptides per protein (mass measurement combined with measurement of the number of cysteines and upstream isoelectric point measurement).

8. References

[1] Kelleher, N. L., Lin, H. Y., Valaskovic, G. A., Aaserud, D. J., Fridriksson, E. K., McLafferty, F. W., Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry, *J. Am. Chem. Soc.*, **1999**, 121(4): 806-812.

[2] McLafferty, F. W., Tandem mass spectrometric analysis of complex biological mixtures, *Int. J. Mass Spectrom.*, **2001**, 212(1-3): 81-87.

[3] Reid, G. E., McLuckey, S. A., 'Top down' protein characterization via tandem mass spectrometry, *J. Mass Spectrom.*, **2002**, 37(7): 663-675.

[4] Meng, F. Y., Du, Y., Miller, L. M., Patrie, S. M., Robinson, D. E., Kelleher, N. L., Molecular-level description of proteins from Saccharomyces cerevisiae using quadrupole FT hybrid mass spectrometry for top down proteomics, *Anal. Chem.*, **2004**, 76(10): 2852-2858.

[5] Patrie, S. M., Charlebois, J. P., Whipple, D., Kelleher, N. L., Hendrickson, C. L., Quinn, J. P., Marshall, A. G., Mukhopadhyay, B., Construction of a hybrid quadrupole/Fourier Transform Ion Cyclotron Resonance Mass Spectrometer for versatile MS/MS above 10 kDa, *J. Am. Soc. Mass Spectrom.*, **2004**, 15(7): 1099-1108.

[6] Patrie, S. M., Robinson, D. E., Meng, F. Y., Du, Y., Kelleher, N. L., Strategies for automating top-down protein analysis with Q-FTICR MS, *Int. J. Mass Spectrom.*, **2004**, 234(1-3): 175-184.

[7] Wu, S. L., Jardine, I., Hancock, W. S., Karger, B. L., A new and sensitive on-line liquid chromatography/mass spectrometric approach for top-down protein analysis: the comprehensive analysis of human growth hormone in an E. coli lysate using a hybrid linear ion trap/Fourier transform ion cyclotron resonance mass spectrometer, *Rapid Commun. Mass Spectrom.*, **2004**, 18(19): 2201-2207.

[8] Finehout, E. J., Cantor, J. R., Lee, K. H., Kinetic characterization of sequencing grade modified trypsin, *Proteomics*, **2005**, 5(9): 2319-2321.

[9] Henzel, W. J., Watanabe, C., Stults, J. T., Protein identification: The origins of peptide mass fingerprinting, *J. Am. Soc. Mass Spectrom.*, **2003**, 14(9): 931-942.

[10] Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., Yates, J. R., Direct analysis of protein complexes using mass spectrometry, *Nat. Biotechnol.*, **1999**, 17(7): 676-682.

[11] Wolters, D. A., Washburn, M. P., Yates, J. R., An automated multidimensional protein identification technology for shotgun proteomics, *Anal. Chem.*, **2001**, 73(23): 5683-5690.

[12] Washburn, M. P., Wolters, D., Yates, J. R., Large-scale analysis of the yeast proteome by multidimensional protein identification technology, *Nat. Biotechnol.*, **2001**, 19(3): 242-247.

[13] Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., Aebersold, R., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags, *Nat. Biotechnol.*, **1999**, 17(10): 994-999.

[14] Smolka, M. B., Zhou, H. L., Purkayastha, S., Aebersold, R., Optimization of the isotope-coded affinity tag-labeling procedure for quantitative proteome analysis, *Anal. Biochem.*, **2001**, 297(1): 25-31.

[15] Han, D. K., Eng, J., Zhou, H. L., Aebersold, R., Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry, *Nat. Biotechnol.*, **2001**, 19(10): 946-951.

[16] Conrads, T. P., Anderson, G. A., Veenstra, T. D., Pasa-Tolic, L., Smith, R. D., Utility of accurate mass tags for proteome-wide protein identification, *Anal. Chem.*, **2000**, 72(14): 3349-3354.

[17] Smith, R. D., Pasa-Tolic, L., Lipton, M. S., Jensen, P. K., Anderson, G. A., Shen, Y. F., Conrads, T. P., Udseth, H. R., Harkewicz, R., Belov, M. E., Masselon, C., Veenstra, T. D., Rapid quantitative measurements of proteomes by Fourier transform ion cyclotron resonance mass spectrometry, *Electrophoresis*, **2001**, 22(9): 1652-1668.

[18] Shen, Y. F., Zhao, R., Belov, M. E., Conrads, T. P., Anderson, G. A., Tang, K. Q., Pasa-Tolic, L., Veenstra, T. D., Lipton, M. S., Udseth, H. R., Smith, R. D., Packed capillary reversed-phase liquid chromatography with high-performance electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry for proteomics, *Anal. Chem.*, **2001**, 73(8): 1766-1775.

[19] Smith, R. D., Anderson, G. A., Lipton, M. S., Pasa-Tolic, L., Shen, Y. F., Conrads, T. P., Veenstra, T. D., Udseth, H. R., An accurate mass tag strategy for quantitative and high- throughput proteome measurements, *Proteomics*, **2002**, 2(5): 513-523.

[20] Lipton, M. S., Pasa-Tolic, L., Anderson, G. A., Anderson, D. J., Auberry, D. L., Battista, K. R., Daly, M. J., Fredrickson, J., Hixson, K. K., Kostandarithes, H., Masselon, C., Markillie, L. M., Moore, R. J., Romine, M. F., Shen, Y. F., *et al.*, Global analysis of the Deinococcus radiodurans proteome by using accurate mass tags, *Proc. Natl. Acad. Sci. U. S. A.*, **2002**, 99(17): 11049-11054.

[21] Petritis, K., Kangas, L. J., Ferguson, P. L., Anderson, G. A., Pasa-Tolic, L., Lipton, M. S., Auberry, K. J., Strittmatter, E. F., Shen, Y., Zhao, R., Smith, R. D., Use of artificial neural networks for the accurate prediction of peptide liquid chromatography elution times in proteome analyses, *Anal. Chem.*, **2003**, 75(5): 1039-48.

[22] Strittmatter, E. F., Ferguson, P. L., Tang, K. Q., Smith, R. D., Proteome analyses using accurate mass and elution time peptide tags with capillary LC time-of-flight mass spectrometry, *J. Am. Soc. Mass Spectrom.*, **2003**, 14(9): 980-991.

[23] Norbeck, A. D., Monroe, M. E., Adkins, J. N., Anderson, K. K., Daly, D. S., Smith, R. D., The utility of accurate mass and LC elution time information in the analysis of complex proteomes, *J. Am. Soc. Mass Spectrom.*, **2005**, 16(1239-1249.

[24] Cargile, B. J., Stephenson, J. L., An alternative to tandem mass spectrometry: Isoelectric point and accurate mass for the identification of peptides, *Anal. Chem.*, **2004**, 76(2): 267-275.

[25] Cargile, B. J., Bundy, J. L., Freeman, T. W., Stephenson, J. L., Gel based isoelectric focusing of peptides and the utility of isoelectric point in protein identification, *J. Prot. Res.*, **2004**, 3(1): 112-119.

[26] Lam, H. T., Josserand, J., Lion, N., Girault, H. H., Numerical design of Off-Gel electrophoresis, **in preparation**, .

[27] Rohner, T. C., Rossier, J. S., Girault, H. H., On-line electrochemical tagging of cysteines in proteins during nanospray, *Electrochem. Commun.*, **2002**, 4(9): 695-700.

[28] Dayon, L., Roussel, C., Prudent, M., Lion, N., Girault, H. H., On-line counting of cysteine residues in peptides during electrospray ionization by electrogenerated tags and their application to protein identification, *Electrophoresis*, **2005**, 26(1): 238-247.

[29] Rohner, T. C., Josserand, J., Jensen, H., Girault, H. H., Numerical investigation of an electrochemically induced tagging in a nanospray for protein analysis, *Anal. Chem.*, **2003**, 75(9): 2065-2074.

[30] Roussel, C., Rohner, T. C., Jensen, H., Girault, H. H., Mechanistic aspects of on-line electrochemical tagging of free L-cysteine residues during electrospray ionisation for mass spectrometry in protein analysis, *ChemPhysChem*, **2003**, 4(2): 200-206.

[31] Roussel, C., Dayon, L., Jensen, H., Girault, H. H., On-line cysteine modification for protein analysis: new probes for electrochemical tagging nanospray mass spectrometry, *J. Electroanal. Chem.*, **2004**, 570(2): 187-199.

[32] Dayon, L., Roussel, C., Girault, H. H., On-line electrochemical tagging of free cysteines in peptides during nanospray ionisation mass spectrometry: an overview, *Chimia*, **2004**, 58(4): 204-207.

[33] Ros, A., Faupel, M., Mees, H., Oostrum, J. V., Ferrigno, R., Reymond, F., Michel, P., Rossier, J. S., Girault, H. H., Protein purification by Off-Gel electrophoresis, *Proteomics*, **2002**, 2(151-156.

[34] Michel, P. E., Reymond, F., Arnaud, I. L., Josserand, J., Girault, H. H., Rossier, J. S., Protein fractionation in a multicompartment device using Off-Gel isoelectric focusing, *Electrophoresis*, **2003**, 24(1): 3-11.

[35] Heller, M., Michel, P. E., Morier, P., Crettaz, D., Wenz, C., Tissot, J. D., Reymond, F., Rossier, J. S., Two-stage Off-Gel (TM) isoelectric focusing: Protein followed by peptide fractionation and application to proteome analysis of human plasma, *Electrophoresis*, **2005**, 26(6): 1174-1188.

[36] Hu, Q. Z., Noll, R. J., Li, H. Y., Makarov, A., Hardman, M., Cooks, R. G., The Orbitrap: a new mass spectrometer, *J. Mass Spectrom.*, **2005**, 40(4): 430-443.

[37] Weiller, G. F., Caraux, G., Sylvester, N., The modal distribution of protein isoelectric points reflects amino acid properties rather than sequence evolution, *Proteomics*, **2004**, 4(4): 943-949.

[38] Nesvizhskii, A. I., Aebersold, R., Interpretation of shotgun proteomics data: the protein inference problem, *Mol. Cell. Proteomics*, **2005**, in press(.

[39] Rappsilber, J., Mann, M., What does it mean to identify a protein in proteomics?, *Trends Biochem. Sci.*, **2002**, 27(2): 74-78.

Chapter 8. Conclusion and perspectives

In the present work, a polymer microspray interface has been tested and characterised for biological mass spectrometry. The primary objective was to validate its use for the analysis of various biomolecules when interfaced with different electrospray ion sources. It has been successfully applied to the analysis of peptides, proteins and glycoconjugates (both purified and from biological sources) on the in-line electrospray ion source of the LCQ Duo ion trap (Thermo Electron), Z-spray source of the LCT Q-TOF (Micromass), the Apollo source of the Apex FT-ICR (Bruker), and a home-made source of an FT-ICR (Alan Marshall's group, National High Magnetic Field Laboratory, Tallahassee, Florida, USA), both in positive and negative ion mode.

Once demonstrated the versatility of this new device, a thorough investigation demonstrated that it performs equally or better than pulled capillary nanosprays, which are usually considered as the most sensitive interface to deliver samples to a mass spectrometer by electrospray. In particular, the polymer microspray showed comparable sensitivity and stability even in the low flow rate range (150-300 nL/min) but a much wider applicable range of flow rates (up to 5 or 7 μL/min depending on the solvent used); above 1 μL/min, it performs better than in the low flow rate range, both in terms of sensitivity and stability. These better performances have been also observed with biological samples: the use of the polymer microspray to ionise sialylated O-glycopeptides from urine has led to reduced sample degradation within the ion source, and observation of multi-sialylated species that had never been observed in the same sample by any other electrospray technique (including

pulled micro- and nanosprays or microfabricated ESI-Chip™ from Advion). This ability to preserve the integrity of labile moieties during ionisation is of utmost importance in this biomedical context, since these studies aim at mapping glycoconjugates present in urine (or other biological tissues, such as brain) in order to understand deficiencies or alteration in the systemic processing of glycoconjugates in particular diseases (such as glycopeptides and glycoproteins in Schindler disease or gangliosides in primary central nervous system tumours). Alleviating instrumental biases in this context can thus prove to be of utmost biomedical relevance. The rationale for these very good performances of the polymer microspray (which are somehow in contradictions with the "rules of thumb" in electrospray mass spectrometry, which states that the smaller the capillary or microchannel used, the better the performances) is not completely clear. On one part, it is not surprising that the polymer microspray can accommodate high flow rates, because its tip shape may be able to accommodate small and large Taylor cones (potentially bigger than the outlet itself by wetting of the smooth outer surface of the microchip). Its good performances at low flow rate and its better softness are more difficult to explain. One possible explanation is that its particular shape leads to the formation of smaller electrospray droplets compared to a circular outlet of comparable dimensions (and hence to more efficient ionisation). But since even the parent droplets are very small, their observation on the flight is a difficult task. One possible line of investigation would be to indirectly measure the internal energy of ions entering the mass spectrometer; if ions generated from the polymer microspray have lower energies than those generated in the same conditions from pulled micro- or nanosprays, it would first provide a rationale for the observed softer ionisation of sialylated O-glycopeptides, and secondly provide some evidence that the primary electrospray droplets formed are smaller than expected with such a device. However, as there are only very few

groups working on such physical characterisation of electrospray in terms of chemical kinetics, there is a crucial lack of reference works to compare with.

One other direction that has been explored in this work concerns more technical developments of the polymer microsprayer itself. First, a polymeric hydrophobic phase has been integrated on the microchip to perform sample clean-up. Though the approach has been willingly kept very crude technically speaking, it proved efficient for peptides and proteins desalting. It is also possible to get rid of other co-solvents usually encountered in protein sample preparation, such as chaotropic (thiourea and urea) and reducing agents. As a stand-alone device, this sample clean-up microchip can be useful in routine mass spectrometric analysis, just as derivatised pipette tips such as ZipTips™ are. It also has the potential of relatively straightforward automation; one possible integration would be to cut PVDF membranes on which proteins have been electroblotted after gel electrophoresis, and put them on the chip just as was done for sample clean up. On-membrane sample processing (clean-up, reduction, proteolysis, enzymatic processing, chemical tagging…) can then be performed with great versatility before mass spectrometric analysis. One other possible approach could be to use membranes with different functionalities; the recent commercial release of membranes for covalent attachment of proteins or antibodies (Empore Affinity AZ™ membranes from 3M) opens possible applications in protein interaction studies or immunoassays, with the advantage of large capture areas compared to what could be obtained with a solid-phase integrated within the microchannel itself.

The other technical development that has been introduced in close collaboration with Joël Rossier from DiagnoSwiss, is a dual channel microsprayer, that allows to mix two solutions within the Taylor cone, just as electrospray droplets are produced. The primary objective of this new design was to allow the analysis of pure aqueous solutions by integrating an organic sheath flow. This proved to be feasible, with the additional advantage

that the organic sheath flow can be kept at a minimum ratio compared to the aqueous sample, so that analysis of proteins in their native form is possible. Moreover, as the ratio between the aqueous sample flow rate and the organic sheath flow rate can be easily controlled, protein conformational studies become feasible by online protein denaturation by organic or acid modifiers, within the Taylor cone. One possible other route is to heat the microchip externally in order to observe heat-induced protein denaturation. Obviously there is still a lot of characterisation to be done, especially given the fact that there is not a definitive consensus about the correlation between charge state distributions observed in mass spectra and the protein conformation state. This lack of consensus is largely due to the fact that there are relatively little data available in the literature (and mostly about model proteins) because such experiments are tedious and difficult to perform with current technologies. If the dual channel microsprayer is improved in terms of robustness and ease of use, which seems at hand, it would enable more comprehensive studies about conformational effects; in this respect, the hyphenation of an ion mobility analyser with a time-of-flight analyser would allow to measure protein collisional cross sections and charge state distributions at the same time, which would allow to confirm or infirm if charge state distributions in ESI-MS truly reflects the conformation of proteins. If so, ESI-MS could become a workhorse for protein thermodynamic characterisation, and a first tool in high-throughput protein structure determination projects.

The last part of this work was devoted to the *in silico* evaluation of the integration of different analytical tools developed in the laboratory in the last three years: on hone hand, Off-Gel electrophoresis allows the separation of complex proteins and peptides mixtures and the measurement of their isoelectric point with very good resolution (down to 0.1 pH unit in routine). Off-Gel electrophoresis combines the power of isoelectric fractionation, such as high resolution and large preconcentration effect) with high loading capacity in

reduced volumes of the order of 0.5-3 mL. On the other hand, a procedure for online counting of cysteine residues within peptides by mass spectrometry without any MS/MS has been developed. The goal of this last part was thus to evaluate how these technologies could perform in the context of high-throughput proteome profiling. There is indeed a clear need to speed-up proteome analysis to enable clinical studies or even clinical use of proteomic technologies, which are still relatively slow and complex to operate. One possible route could be to seek for proteolytic peptides that are unique (in terms of their measured physical properties) within the whole digested proteome. If peptide mass only is measured, even ultrahigh resolution and mass accuracy mass spectrometers can hardly provide useful information, because there are very few unique peptides in terms of mass in whole digested proteomes; on the contrary, if peptides isoelectric point is measured, as well as the number of cysteines within each peptide, there are much more unique peptides, and fast proteome profiling can be envisaged even in complex proteomes such as human proteomes.

From a more general point of view, it is interesting to put this work in perspective with the development of proteomics itself; it indeed started from relatively technical work, namely the development of new electrospray interfaces, and evolved in various directions: on one hand the development of relatively simple tools such as the online SPE microspray, or the dual microsprayer with the initial idea to ease the analysis of aqueous solutions… on the other hand the dual microsprayer turned out to open new possibilities for protein structural characterisation, and the integration of different tools that were purely analytical chemistry projects allows to envision an entirely new proteomic platform. Bioinformatics came into play to evaluate the feasibility of such a platform for fast proteome profiling. And indeed the development of proteomics has followed such a trend, from very analytical work to build the analytical tools to study protein on a larger and larger scale. The availability of

separation methods and mass spectrometers has progressively allowed researchers to imagine an entirely new scale of research, which in turn has generated needs for new instruments and methodologies. Whereas it is certain that progress will still be made on the instrumental side, and in particular that high resolution mass spectrometers will become more and more affordable, one can hypothesise that methodology development allied with more robust data analysis tools will drag the field of proteomics in the very next few years. This also implies that further developments in the field will require larger and larger facilities, and more integrated work between instrument developers, database developers, software engineers and statisticians. It is not certain yet how much of new biological knowledge will come out from the massive efforts undertaken worldwide; when data robustness is acquired, biological modelling will have to play a role to interpret such massive data, and extract some relevant information for diagnosis, prognosis, treatment follow-up and drug design, and this will consist in a new revival for theoretical biology. How much can we hope from this new revolution in biology? It will probably take no less than a few decades to find the answer.

Niels Lion (French nationality)
Birth: May 1974, 24th
Physical and Analytical Electrochemistry Lab
Institute of Chemical Sciences and Engineering
Ecole Polytechnique Fédérale de Lausanne
1015 Lausanne
Switzerland
+41-(0)21-693-36-50
niels.lion@epfl.ch
http://lepa.epfl.ch

Graduate research assistant in micro-analytical sciences for proteomics

Skills

Proteomics, protein biochemistry and analytical biochemistry:
Mass spectrometry (ESI-IT-MS and ESI-TOF-MS) for protein analysis, classical techniques in analytical biochemistry (UV and fluorescence spectroscopies, CD, chromatographies, gel and capillary electrophoresis…), molecular biology (PCR, sequencing, molecular engineering, mutants rational design…)

Instrumentation:
Polymer microtechnologies, microfluidics, signal and image acquisition and processing (C programming language, LabView, MatLab…), fluorescence instrumentation (microscopy, evanescent waves…).

Others:
Review of manuscripts submitted to Electrophoresis, Rapid Communications in Mass Spectrometry, Journal of the American Society for Mass Spectrometry, Journal of Chromatography A… Tight involvement in the writing of proposals and management of scientific projects at the European and Swiss level. Management of several trainees at different levels, on analytical instrumentation (BSc level, Integration of a desalting step into MS microchips), image processing (MSc level, 3D microparticles counting and tracking in confocal microscope images), and microbiology (technician level, development of a screen for heavy metal resistance for *E. Coli* strains expressing recombinant proteins).

Professional experiences

July 2001-January 2006:	Ph.D. on the development on **microfluidic front-end technologies for electrospray mass spectrometry**. *Physical and Analytical Electrochemistry Lab. Institute of Chemical Sciences and Engineering. Swiss Federal Institute of Technology in Lausanne (EPFL). PhD supervisor: Prof. H. H. Girault.*
July 2001-present:	**Teaching assistant** for bachelor and master students in chemistry (*EPFL and Lausanne University*): general chemistry (practices: 384 h.) and separation sciences (exercise sessions: 112 h.). **Award for teaching excellence in 2004.**
Sept. 1999-June 2001:	Development of **fluorescence instrumentation for DNA microarrays** in the framework of a DNA sequencing project. *BioMérieux and Commissariat à l'Energie Atomique. Laboratoire d'Electronique, de Technologie et d'Instrumentation (LETI). Grenoble. France*
Nov. 1998-Aug. 1999:	**Molecular engineering of the Green Fluorescent Protein** to develop a metal biosensor. Military service. *Commissariat à l'Energie Atomique. Saclay. France.*

March-Aug. 1998:	Development of a data compression algorithm for **cardiac implantable devices** to allow patient management. *Ela Recherche. Le-Plessis-Robinson. France*
April 1998:	Bibliographic project on electron-beam computerized tomography. *Ecole Centrale de Paris. France*.
June-Aug. 1997:	Feasibility study of a **sudden-infant-death-syndrome (SIDS) diagnostic procedure** by spectral analysis of heart rate variability. *Institute of Medical Technology and Neonatalogy Intensive Care Unit. University and Regional Hospital. Lille. France.*

Publications in preparation

1. **Hyphenation of microfluidic devices with electrospray mass spectrometry,** Niels Lion, Joël S. Rossier, Franta Foret, Hubert H. Girault (review article).
2. **Evaluation of mass spectrometry front-end technologies to increase peptide uniqueness in proteomics,** Niels Lion, Joël S. Rossier, Hubert H. Girault.
3. **New method for detection of Proteins on PVDF membranes and PET microchannels using copper staining combined with Scanning Electrochemical Microscopy,** Maurizio Carano, Niels Lion, Hubert H. Girault.
4. **Numerical design of an isoelectric focusing device,** Hoang-Trang Lam, Jacques Josserand, Niels Lion, Hubert H. Girault.
5. **Combined chip-based separation and MS, instrumentation and application,** Niels Lion, Tatiana C. Rohner, Laura Bindila, Martin Froesch, Alina Zamfir, Jasna Peter-Katalinic, Hubert H. Girault, *Encyclopedia of Mass Spectrometry*, 2006, vol. 8: Hyphenated methods, principles and instrumentation, Wilfried Niessen Ed., Elsevier.

First author publications in peer reviewed journals

1. **Editorial: microfluidics in systems biology**, Niels Lion, Joël S. Rossier, Hubert H. Girault, 2005, *Electrophoresis,* 26: 3593.
2. **Thin-chip microsprayer system coupled to quadrupole time-of-flight mass spectrometer for glycoconjugate analysis**, Alina Zamfir*, Niels Lion*, Zeljka Vukelic, Laura Bindila, Joël S. Rossier, Hubert H. Girault, Jasna Peter-Katalinic, 2005, *Lab on a Chip*, 5: 298-307. (* equal contribution)
3. **Thin-chip microsprayer system coupled to Fourier Transform Ion Cyclotron Resonance mass spectrometry for glycoconjugate analysis**, Laura Bindila*, Martin Froesch*, Niels Lion*, Zeljka Vukelic, Joël S. Rossier, Hubert H. Girault, Jasna Peter-Katalinic, Alina Zamfir, 2004, *Rapid Communications in Mass Spectrometry*, 18: 2913-2920. (* equal contribution)
4. **Flow-rate characterization of microfabricated polymer microspray emitters**, N. Lion, J.O. Gellon, H. H. Girault, 2004, *Rapid Communications in Mass Spectrometry*, 18(14): 1614-1620.
5. **Why the move to microfluidics for protein analysis?**, N. Lion, F. Reymond, H.H. Girault, J.S. Rossier, 2004, *Current Opinion in Biotechnology*,.15(1): 31-37.
6. **Microfluidic systems in proteomics** (review article), N. Lion, T. C. Rohner, L. Dayon, I. L. Arnaud, E. Damoc, N. Youhnovski, C. Roussel, J. Josserand, H. Jensen, J. S. Rossier, M. Przybylski, H. H. Girault, 2003, *Electrophoresis*, 24(21): 3533-3562.
7. **On-chip protein sample desalting and preparation for direct coupling with electrospray ionization mass spectrometry**, N. Lion, J.O. Gellon, H. Jensen, H.H. Girault, 2003, *Journal of Chromatography A*, 2003, 1003(1-2): 11-19.
8. **Integration of a membrane-based desalting step in a microfabricated disposable polymer injector for mass spectrometric protein analysis**, N. Lion, V. Gobry, H. Jensen, J.S. Rossier, H.H. Girault, 2002, *Electrophoresis*, 23(20): 3583-3588.

Publications in peer reviewed journals

1. **Chip electrospray mass spectrometry for carbohydrate analysis**, Alina Zamfir, Laura Bindila, Niels Lion, Mark Allen, Hubert H. Girault, Jasna Peter-Katalinic, 2005, *Electrophoresis*, 26: 3650-3673.
2. **SPS' Digest: The Swiss Proteomics Society selection of proteomics articles**, Christine Hoogland, Niels Lion, Jean-Charles Sanchez, Jean-Daniel Tissot, 2005, *Proteomics.* 5: 3045–3047.
3. **Copper staining/labeling and Scanning Electrochemical Microscopy readout of proteins on poly(vinylidene difluoride) membranes,** Maurizio Carano, Niels Lion, Hubert H. Girault, 2005, *Chimia*, 59(3): 105-108.
4. **Salt removal during Off-Gel electrophoresis of protein samples,** Isabelle Arnaud, Jacques Josserand, Henrik Jensen, Niels Lion, Christophe Roussel, Hubert H. Girault, 2005, *Electrophoresis,* 26: 1650-1658.

5. **On-line counting of cysteine residues in peptides during electrospray ionization by electrogenerated tags and its application to protein identification**, Loïc Dayon, Christophe Roussel, Michel Prudent, Niels Lion, Hubert H. Girault, 2005, *Electrophoresis*, 26: 238-247.
6. **Generation of mass tags by the inherent electrochemistry of electrospray for protein mass spectrometry**, Christophe Roussel, Loïc Dayon, Niels Lion, Tatiana C. Rohner, Jacques Josserand, Joël S. Rossier, Henrik Jensen, Hubert H. Girault, 2004, *Journal of the American Society for Mass Spectrometry*, 15: 1767-1779. (review article)
7. **On-line electrochemical tagging of free cysteines during nanospray ionisation for mass spectrometry analysis,** Christophe Roussel, Loïc Dayon, Tatiana C. Rohner, Niels Lion, Hubert H. Girault, 2004, *Progress in Biomedical Optics and Imaging*, 5(32): 30-37.
8. **Detection of proteins on poly(vinylidene difluoride) membranes by scanning electrochemical microscopy**, Maurizio Carano, Niels Lion, Jean-Pierre Abid, Hubert H. Girault, 2004, *Electrochemistry Communications*, 6: 1217-1221.
9. **Electrochemical and theoretical aspects of electrospray ionization** (invited article), T. C. Rohner, N. Lion, H. H. Girault, 2004, *Physical Chemistry Chemical Physics*, 6(12): 3056-3068. (selected as Royal Society of Chemistry hot article)
10. **Thin-chip Microspray system for high-performance Fourier-transform ion cyclotron resonance mass spectrometry of biopolymers**, J.S. Rossier, N. Youhnovski, N. Lion, E. Damoc, F. Reymond, H.H. Girault, M. Przybylski, 2003, *Angewandte Chemie-International Edition*, 42(1): 53-58.

Patent

1. **Sample dispensing apparatus for mass spectrometry analysis, includes polymer substrate portion, defining portion of microchannel wall, having electrically or ionically conductive material to apply voltage to the sample,** Joël S. Rossier, Frédéric Reymond, Hubert H. Girault, Tatiana Rohner, Niels Lion, Véronique Gobry, patent numbers: WO200280222-A, EP1366506-A, WO200280222-A1, EP1366506-A1, US2004075050-A1, AU2002302369-A1.

Non peer reviewed publications

1. **Tagging of peptides during nano electrospray ionisation mass spectrometry**, Christophe Roussel, Loïc Dayon, Tatiana C. Rohner, Niels Lion, Hubert H. Girault, 2004, *Bioforum Europe*, 4: 40-43
2. **Tagging of peptides during nano electrospray ionisation mass spectrometry**, Christophe Roussel, Loïc Dayon, Tatiana C. Rohner, Niels Lion, Hubert H. Girault, 2004, *GIT Laboratory Journal Europe*, 3: 41-43.
3. **Marquage de peptides lors d'une spectrométrie de masse nano electrospray,** Christophe Roussel, Loïc Dayon, Tatiana C. Rohner, Niels Lion, Hubert H. Girault, 2004, *Bioforum France*, 2: 30-32.
4. **Nouveau défi pour l'électrophorèse**, Niels Lion, Hubert H. Girault, 2003, *Tracés*, 23 :14-15.
5. **La microfluidique,** Niels Lion, Hubert H. Girault, 2003, *Tracés*, 23 : 16-17.

Conferences and courses

- 1er Symposium de Chimie et Biologie Analytique, Montpellier, September 2005, 26-29, **invited oral presentation: "On-line electrochemical tagging for proteomics"** (Niels Lion, Hubert H. Girault)
- CESAGen 2nd International Conference "Genomics and society", London, April 2005, 12-14, **Oral presentation:** "Underestimated contributions in the conceptual advent of systems biology: a (naive) scientist point of view" (Niels Lion)
- Swiss Proteomic Society annual meeting: Proteomics for health: from development to applications, Bern, December 6-8, 2004. **Poster presenter**: "Thin chip polymer microsprays for glycoconjugate analysis of Schindler disease patients" (Niels Lion, Laura Bindila, Martin Froesch, Zeljka Vukelic, Joël S. Rossier, Hubert H. Girault, Jasna Peter-Katalinic, Alina Zamfir)
- Swiss Proteomic Society annual meeting: Proteomics for health: from development to applications, Bern, December 6-8, 2004. **Poster presenter**: "Detection of proteins on poly(vinylidene difluoride) membranes by scanning electrochemical microscopy" (Maurizio Carano, Niels Lion, Hubert H. Girault)
- Swiss Chemical Society Fall meeting 2004, Zurich, October 7th, 2004. **Oral presentation**: "Microfluidic front-end technologies for electrospray mass spectrometry" (Niels Lion, Loïc Dayon, Christophe Roussel, Hubert H. Girault)
- French Society for Mass Spectrometry 21st annual meeting, Strasbourg, September 14-17, 2004. **Oral presentation**: "Microfluidic front-end technologies for electrospray mass spectrometry" (Niels Lion, Loïc Dayon, Christophe Roussel, Hubert H. Girault)
- 10th International Conference on Electroanalysis, European Society for ElectroAnalytical Chemistry, National University of Ireland, Galway, 6-10 June 2004. "Scanning electrochemical microspcopy for on-membrane detection of low abundant proteins" (Carano M, Girault HH, Lion N) **Best poster award**.

- Swiss Proteomic Society Annual Meeting", Basel (Switzerland), December 2003, **Poster presenter**: "Creation of a virtual proteomics journal" (N. Lion, C. Hoogland, D. Noukakis, P. Palagi, F. Muller, R. Stöcklin, J.C. Sanchez, J.S. Rossier, J.D. Tissot)
- Swiss Proteomic Society Annual Meeting", Basel (Switzerland), December 2003, **Poster presenter**: "Characterization of native proteins by ESI-MS thanks to a dual-channel microfabricated electrospray emitter" (N. Lion, J.S. Rossier, F. Reymond, H. H. Girault)
- Swiss Proteomic Society Annual Meeting", Basel (Switzerland), December 2003, Poster: "On-line electrochemical tagging of free cysteine by electrogenerated benzoquinones during nanospray ionisation for Mass Spectrometry for proteomics. Part 2" (L. Dayon, C. Roussel, N. Lion, M. Prudent, T.C. Rohner, H. Jensen and H.H. Girault)
- "20th Montreux Symposium on liquid chromatography / mass spectrometry", October 2003, Savannah, Georgia, USA. **Oral communication**: "Characterization of native proteins by ESI-MS thanks to a dual-channel microfabricated electrospray emitter" (N. Lion, J. S. Rossier, F. Reymond, H. H. Girault).
- " Mass spectrometry and Proteomics", September 2003, Toulouse (France), joint congress of the French Mass Spectrometry Society and French Society for Electrophoresis and Proteome Analysis. **Oral communication**: "Polymer nanospray chips for proteome analysis".
- "Introduction to proteomics: Investigation fields and methodologies", May 2003, Marseille (France), Workshop organized by the Centre National de la Recherche Scientifique, **Invited lecture**: The Use of Microsystems in proteomics (N. Lion, J.S. Rossier, H. H. Girault).
- 5th Igler MS Tage, Innsbruck, Austria, February 2003. **Poster presenter**: On-chip protein sample preparation for direct coupling with ESI-MS (N. Lion, J.O. Gellon, V. Gobry, H.H. Girault).
- 2nd Annual Conference of the Swiss Proteomic Society, Applied Proteomics, Lausanne, 2002, 3-5 December. **Poster presenter**: Integration of a desalting step into polymer microchips: applicability to proteomic sample preparation (N. Lion, V. Gobry, J.S. Rossier, H.H. Girault).
- 2nd Annual Conference of the Swiss Proteomic Society, Applied Proteomics, Lausanne, 2002, 3-5 December. **Poster presenter**: Automated disposable nanospray chips interfaced to ion trap and Q-trap mass spectrometers (J.S. Rossier, V. Gobry, N. Lion, G. Lagger, H.H. Girault, E. Varesio, L.A. Leuthold, G. Hopfgartner, F. Reymond)
- 19th Symposium on Liquid Chromatography/ Mass Spectrometry, Montreux, 2002, 4-8 November. **Poster presenter**: Integration of a membrane-based desalting step in a MS-microchip (N. Lion, V. Gobry, J.S. Rossier, H.H. Girault).
- High-resolution Biopolymer Mass Spectrometry International Course, German Chemical Society, Konstanz University, 22-26 October 2001.
- First meeting of the Swiss Proteomics Society (Functional Proteomics, Geneva, 21-22 November 2001). **Poster presenter**: MS microchip for proteomics (N. Lion, T. Rohner, V. Gobry, J. Rossier, H.H. Girault).
- Microarrays 2 Macroresults, Advancing drug development, Boston, April 23-25, 2001. **Poster presenter**: Direct fluorescence imaging system for melting studies on DNA microarrays (N. Lion, F. Perraut, A. Guiot, P. Puget, F. Ginot).

Education

- Master of Science with honors in medical and biological engineering (signal and image processing). October 1998. University Paris XII-Val de Marne
- Master of Science with honors in Industrial Biotechnologies. September 1998. Ecole Centrale de Lille. France.
- English: Certificate in Advanced English, grade B, Cambridge.
- German: beginner.

Personal

- Editor of a special issue of Electrophoresis entitled "Microfluidics in system biology" (September 2005)
- Member of the Swiss Chemical Society (division of analytical chemistry)
- Committee member of the Swiss Proteomic Society (initiator of the SPS Digest, a virtual journal of proteomics)
- Member of the Swiss Group for Mass Spectrometry.
- Personal interest in sociology and philosophy of science and technology.